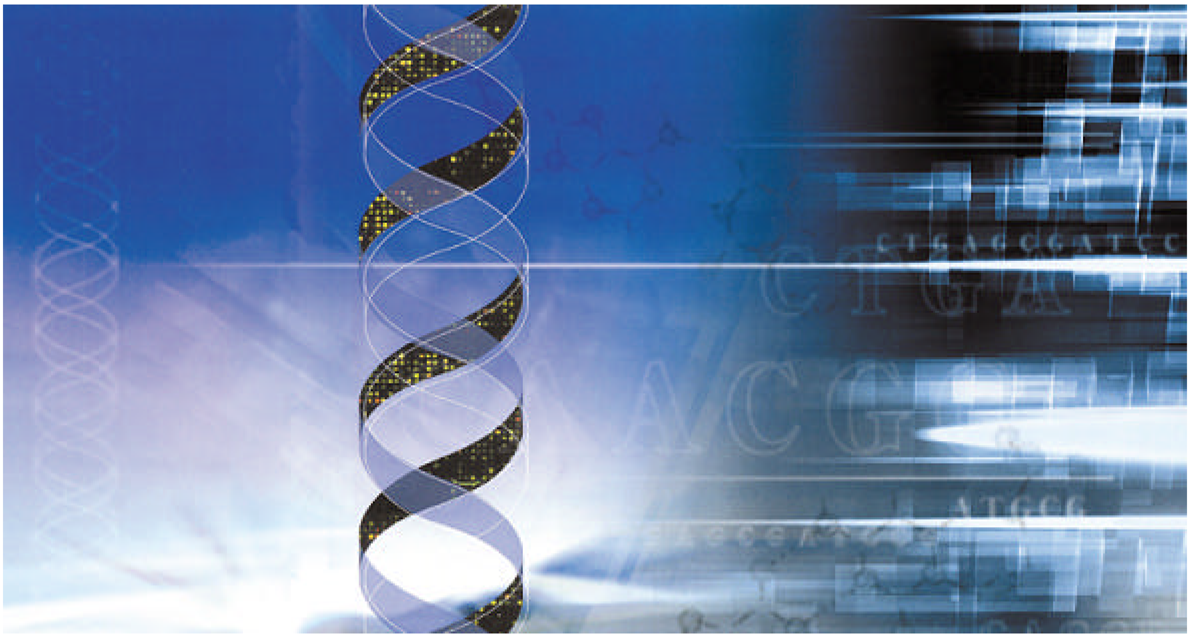
Microarray Data Analysis Tool

Manual Ver3.0　2008 / 3



お　願　い

本ソフトウェアの著作権はフィルジェン(株)に帰属します。ソフトウェアの全部または一部を無断で複製したり、無断で複製物を頒配すると、著作権の侵害になりますのでご注意下さい。また、本ソフトウェアはフィルジェン(株)の受託解析データ専用のものであり、弊社以外のデータを用いて使用することを一切禁じます。

本ソフトウェアをご使用前には本取扱説明書を必ずお読み下さい。なお、本取扱説明書は大切に保管して下さい。また、本ソフトウェアまたは取扱説明書の内容に関して、改良のため、将来予告なしに変更することがあります。

フィルジェン株式会社

バイオサイエンス受託解析センター
〒456-0032　名古屋市緑区大高町中ノ島 15 番地 1
Tel　：（052）-624-4388　(9:00～17:00)
Fax　：（052）-624-4389
E-Mail　：biosupport@filgen.jp

http://www.filgen.jp

目 次

1.	ソフトウェア概要	P.2
2.	ソフトウェア起動と終了	P.3
3.	ソフトウェアの操作（メイン画面）	P.4～17
4.	ソフトウェアの操作（検索：Filter option）	P.18～28
5.	ソフトウェアの操作（Scatter Plot）	P.29～32
6.	ソフトウェアの操作（統計検定）	P.33～34
7.	ソフトウェアの操作（Gene Ontology 解析）	P.35～38
8.	ソフトウェアの操作（Pathway 解析）	P.39～41
9.	ソフトウェアの操作（External Filter 機能）	P.42～43
10.	ソフトウェアの操作（クラスタリング用データ出力）	P.44
11.	ソフトウェアの操作（Cluster and TreeView 簡易マニュアル）	P.45～46
12.	よくあるご質問	P.47～48
13.	履歴	P.49

1. 概 要

本ソフトウェアは、弊社の受託解析データ専用に開発されたマイクロアレイ用データ解析ソフトです。
膨大な解析データから希望の結果を抽出するための、各種フィルタリングや検索機能や公共 Database 等へハイパーリンク機能を搭載しています。また、自分の興味ある遺伝子（or 抗体）群を登録できる遺伝子（or 抗体）セット登録機能やスキャッタープロットをはじめとするグラフ作成機能、グループ間の発現量に有意な差があるかを調べる統計検定機能、Gene Ontology、Pathway による遺伝子機能による分類も行えます。さらに、各データは Stanford 大学 Eisen,Lab からダウンロードできる統計解析フリーソフトウェア「Cluster and TreeView」のファイルフォーマットに出力させることができます。

【機能一覧】
- Ensembl,RefSeq,Gene Ontology をはじめとする公共データベース等の ID やキーワードによる検索。
- Ensembl,RefSeq,Gene symbol をはじめとする公共データベース等へのハイパーリンク機能。
- 染色体番号別検索。
- カテゴリー別分類検索（Human,Mouse,Rat に対応）。
  Apoptosis ･Cancer ･Cell Cycle ･Cytokine ･Common Diseases ･Extracellular Matrix and Adhesion Molecules ･Neuroscience ･Signal Transduction ･Stem Cell ･Toxicology and Drug Metabolism etc
- 発現 Ratio や Cutoff 値などの条件を入力して検索するフィルタリング機能。
- 複数実験検索機能。2 つ以上実験において共通に変動する遺伝子（or 抗体）、あるいは実験特異的に変動している遺伝子（or 抗体）などの実験群に対するデータの検索も可能。
- 単色多サンプルデータに対するデータの加工･検索機能。各種ノーマライズ、サンプル間の発現比計算、サンプルデータのグループ間検索機能を搭載。
- 遺伝子（or 抗体）セット登録機能。
- スキャッタープロット作成機能。各プロットをクリックするとプロット専用画面が表示。
- 実験データのグループ分け。
- グループ分けを行った実験データ群に対するグループ別実験条件検索。
- 複数のグループ間における発現量の統計解析。
- Gene Ontology、GO Slim、GenMAPP Pathway の階層構造表示。
- Gene Ontology、GO Slim、GenMAPP Pathway の出現頻度による、サンプルの生物学的機能解析。
- 指定のアノテーション項目を基準にしたデータの平均化。
- 自分で作成したデータリストと、ソフトウェア中の指定のアノテーション項目のマッチング。これにより、マイクロアレイデータと外部のデータファイルの統合が可能。

＊本ソフトウェアは遺伝子発現解析をベースとして作成されており、ソフトウェア上の表記等も遺伝子発現解析用となっております。抗体アレイなど遺伝子発現解析以外で本ソフトウェアをご利用の場合はソフトウェア上の表記が異なりますのでご了承下さい。
＊使用するマイクロアレイの種類により、付加される公共データベースの情報、一部対応していない機能などがありますが、ご了承下さい。

【動作環境】 以下の OS でご使用下さい。
･Windows Xp
･Windows 2000

＊解析の際に大量のデータを扱うことが予想されるため、Windows98 等（95 系）では OS 固有の問題（リソースの問題）により、グラフ表示等が正常に行われない場合があります。
＊実験データ、アノテーションデータを読み込む際に、それらのファイルをエクセルで開いているとエラーとなりますのでご注意ください。（エクセルから他のアプリケーションでファイルを開けないよう排他ロックをかけられているためです。）
＊本ソフトウェアの起動中は、他のアプリケーションソフトウェアとの同時使用はお避け下さい。コンピュータ自体がハングアップし、本ソフトウェアに悪影響を及ぼし、作成データの保存が出来なくなる場合があるので、ご注意下さい。

2　ソフトウェアの起動と終了

注意

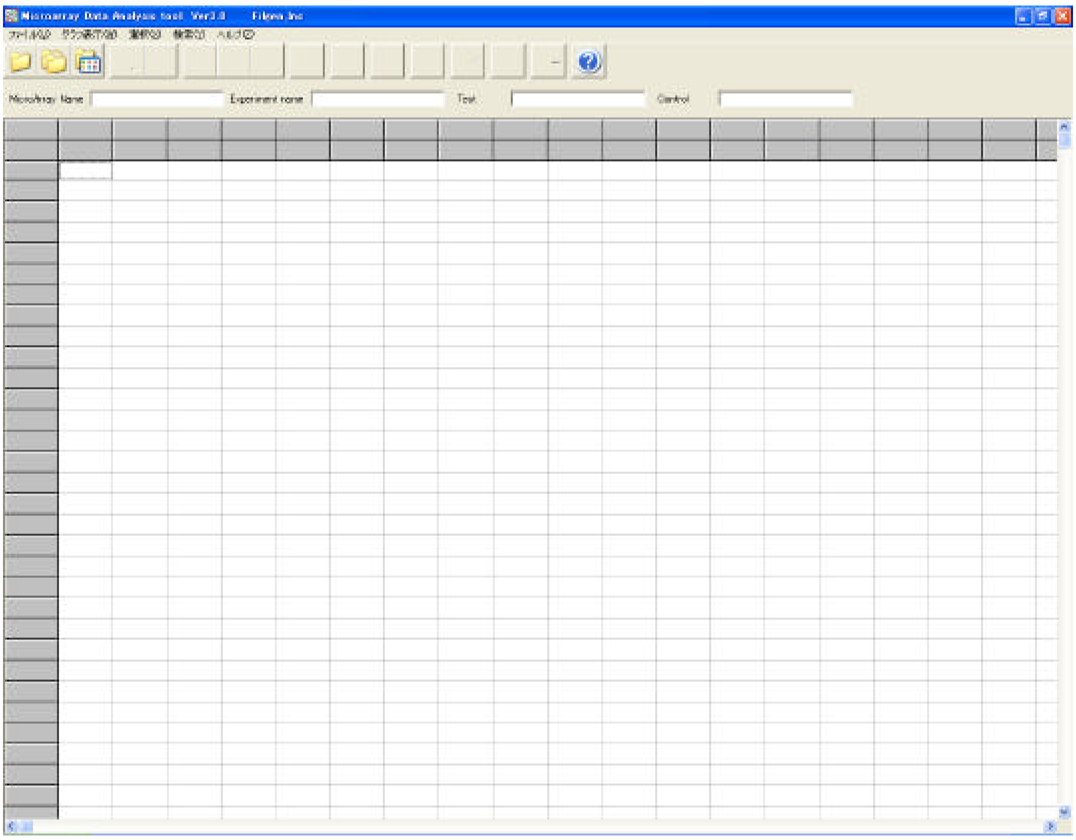
ソフトウェア起動と終了

本ソフトウェアの起動と終了は以下の手順に従って下さい。

1. CD-ROM「受託解析サービス実験データ」内にあるフォルダ「MicroarrayDataAnalysisTool」をフォルダごとコピーし、ご使用になられるパソコンにペーストしてください。ペースト後にフォルダを開きます。CD-ROM から直接ソフトウェアを起動すると、ソフト終了時における設定ファイルの自動保存が行えないため、エラーメッセージが表示されますのでご注意ください。



2. アプリケーションファイル「MAtool.exe」をダブルクリックします。タイトル画面が表示されますので、完全に起動するまでお待ちください。起動後、下図のメイン画面が表示されます。なお、フォルダ「実験データ」には本ソフトウェア専用にフォーマットされた受託解析データがあります。実験データを読み込む際は、このフォルダを指定して、データの読み込みを行ってください。



3. 本ソフトウェアを終了させる場合は、メイン画面のファイル（V）から「終了」を選択する、あるいはメイン画面右上隅の ボタンをクリックします。「終了しますか」という確認メッセージが表示されますので、よければ「はい」を選択し、ソフトウェアを終了させます。

3.　ソフトウェアの操作　（メイン画面）

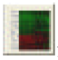
表示アイコンの説明　　メイン画面で表示されている表示アイコンの説明を以下に示します。

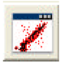
1比較データ解析　：1比較のデータを解析するための解析モードです。

複数比較データ解析　：複数の比較データを同時に解析するための解析モードです。

Signal比較データ解析　：単色データに対して、データ加工および解析を行うための解析モードです。

保存　：実験データをCSVファイル（*.csv），テキストファイル（*.txt）で保存します。
「Select」の項で「○」が付いているものが保存対象となります。

クラスタリング用データ出力　：Stanford大学 Eisen,Labからダウンロードできる統計解析ソフト「Cluster and TreeView」のファイルフォーマット用に表示されているデータをtxtファイルとして出力させます。

Scatter Plot表示　：Scatter Plotグラフの作成を行います。詳細はソフトウェアの操作（Scatter Plot,/Plot Window）をご参照ください。

Plot Window表示　：Scatter Plot上の各プロットの詳細情報を表示します。詳細はソフトウェアの操作（Scatter Plot/Plot Window）をご参照ください。

実験データ表示　：現在選択されている実験データのファイル名およびサンプル名を表示します。

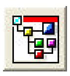
Filter Options　：メイン画面で表示されている実験データに対して各種検索を行います。

グループ編集　：実験データのグループ分けの設定を行います。

平均化　：指定されたアノテーション項目を基準にして、重複したデータを平均化し、1つにまとめます。

統計検定　：グループ間またはグループ内発現量の統計学的検定法の選択を行います。詳細はソフトウェアの操作（統計検定）をご参照ください。

GO解析　：メイン画面で表示されている実験データに対して、Gene Ontology解析を実行します。詳細はソフトウェアの操作（Gene Ontology解析）をご参照ください。

Pathway解析　：メイン画面で表示されている実験データに対して、Pathway解析を実行します。詳細はソフトウェアの操作（Pathway解析）をご参照ください。

External Filter　：実験データと自分で作成した外部のデータを、指定のアノテーション項目を基準にして、統合させます。詳細はソフトウェアの操作（External Filter）をご参照ください。

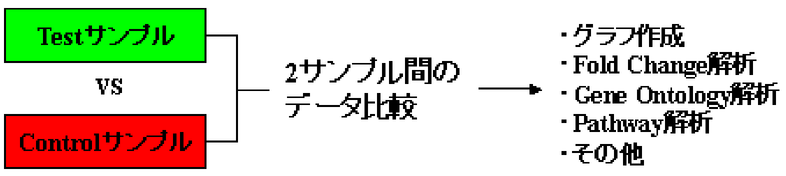
ヘルプ　：本ソフトウェアのマニュアル（PDFファイル）を表示します。

メイン画面

アプリケーションファイル「MAtool.exe」をダブルクリックし、ソフトウェアを起動させると、メイン画面が表示されます。
メイン画面では実験データの読み込みや保存、表示をはじめ本ソフトウェアの基本操作を行います。また、本ソフトウェアでは1比較データ解析、複数比較データ解析、Signal 比較データ解析の 3 つのモードがあります。解析をはじめる前に、どの解析モードを使用するか選択してください。

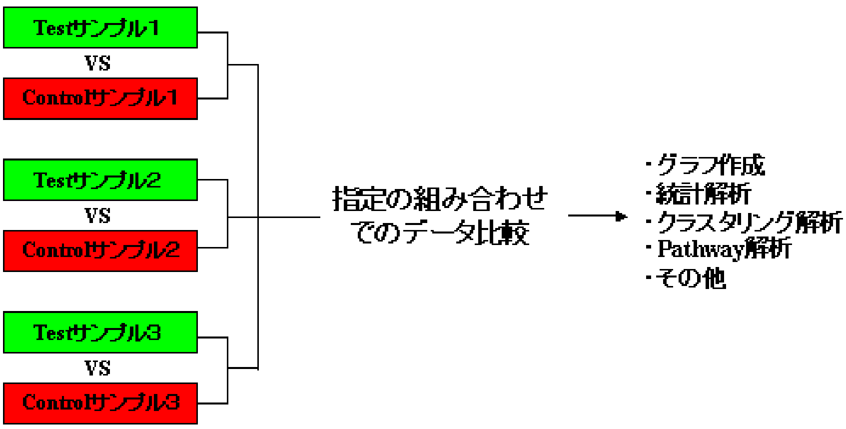
【1比較データ解析】

1実験分の実験データを対象にデータ検索やグラフ作成、機能解析等の解析を行います。



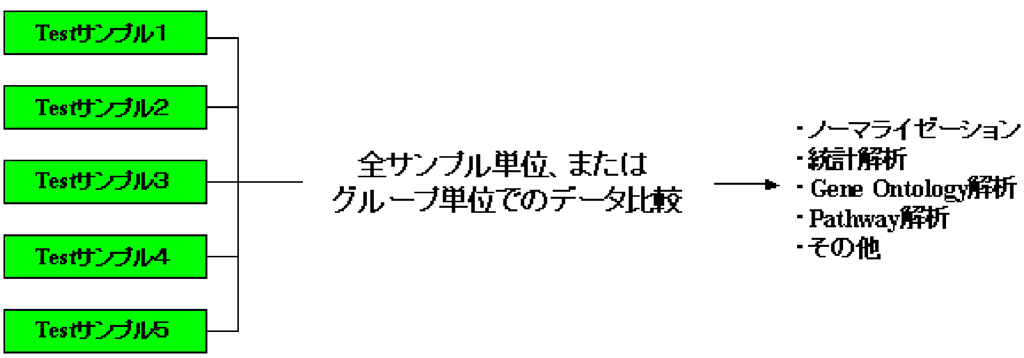
【複数比較データ解析】

2実験以上の実験データを対象にデータ検索やグラフ作成、機能解析等の解析を行います。また複数の実験データをグループ分けし、グループ別データ検索やグループ間発現差の統計検定を行うことができます。



【Signal比較データ解析】

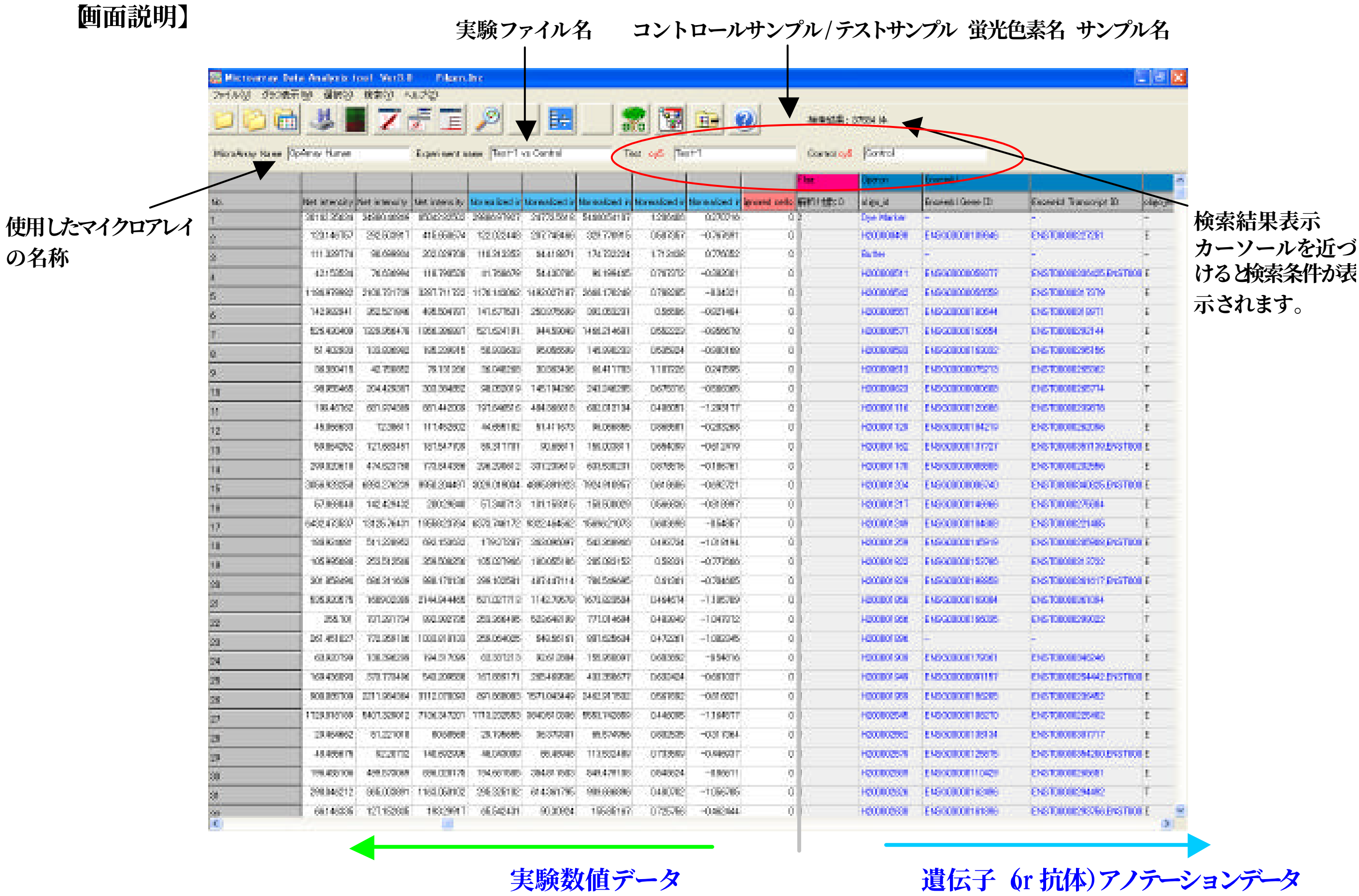
2実験以上の単色実験データを対象にデータ加工・検索やグラフ作成、機能解析等の解析を行います。複数の実験データをまとめてノーマライズし、実験間発現比の計算を行うことができます。また複数比較データ解析と同様、実験データのグループ化や統計検定も行えます。



1比較データ解析

1比較データ解析の場合　：アイコンをクリックし、解析対象となる1比較分の実験データを読み込みます、下図のような実験数値データ(背景：白)及びアノテーション情報(背景：グレー)が表示されます。Ignored cells より左側が実験数値データ、右側が遺伝子（or 抗体）アノテーションデータ情報になります。
＊データ数の多いファイルの場合、読み込み時間が30～60秒程度かかります。

【画面説明】

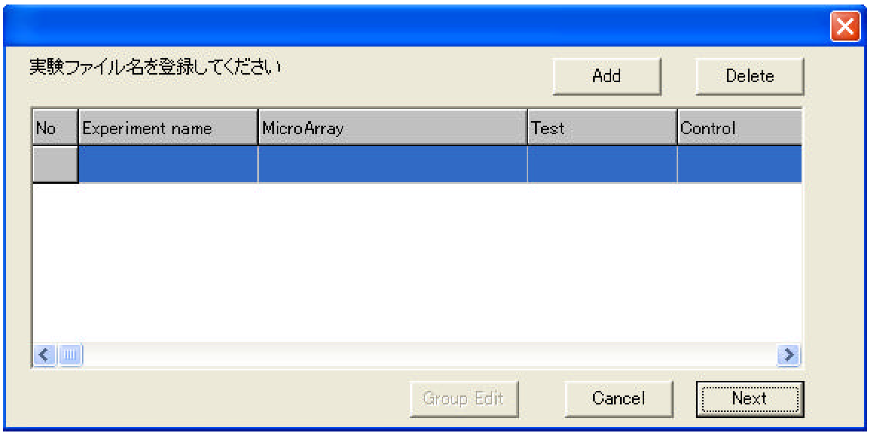


＊この1比較データ解析で表示するデータは、納品Excelデータとまったく同じものです。
ただし納品用Excelデータと違い、変動遺伝子を抽出するための「2Up」、「2Down」フラグがありません。
そのため変動遺伝子を抽出するためには、Filter Options機能(P.18参照)を使用し、数値データをもとに抽出を行う必要があります。詳細な抽出の手順は、「よくあるご質問(P.47)」を参考にしてください。
＊1比較データ解析では、実験データが1つしかないため、データのグループ化と発現量の統計検定が行えません。そのため、「Group Edit」アイコンと「統計検定」アイコンは、選択できないようになっています。

複数比較データ解析

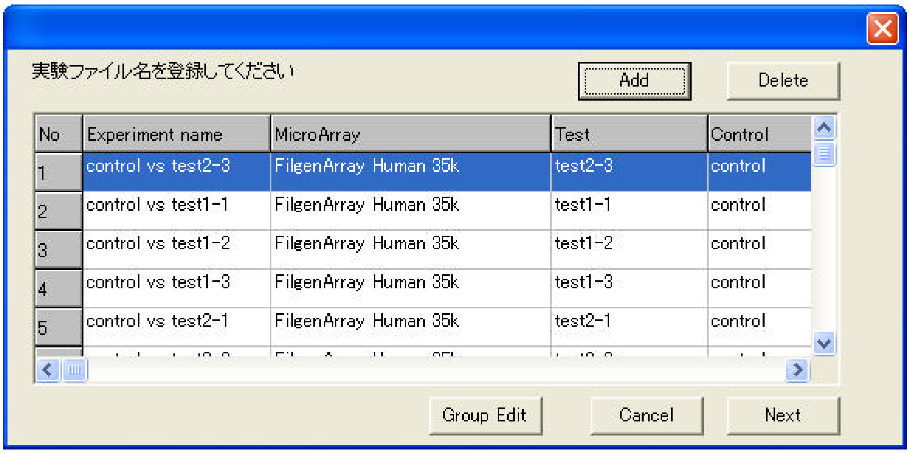
複数比較データ解析の場合　：　2比較以上の実験データを対象にデータ検索等の解析を行います。

①アイコンをクリックすると、下記の画面が表示されます。「Add」ボタンをクリックし、対象となる実験データの選択 読み込みを行います。



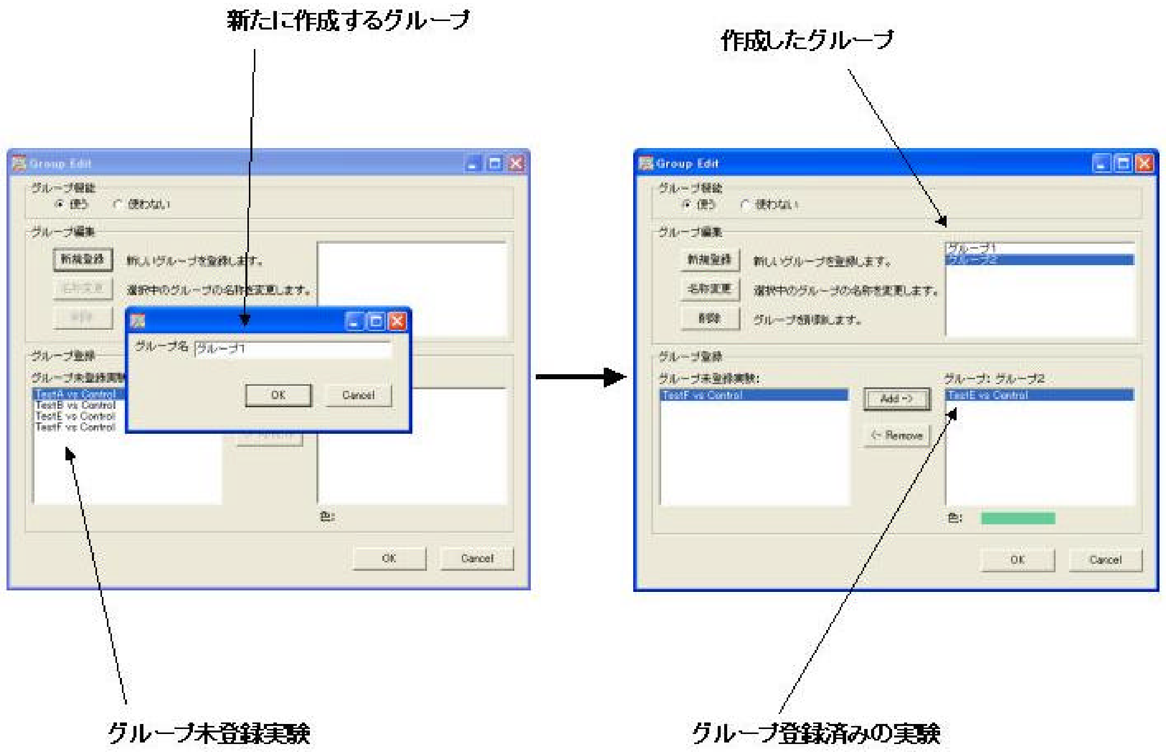
＊　複数比較データ解析で読み込める実験データは同一規格のマイクロアレイに限ります。
＊　実験データ数が多い場合は読み込み時間が長くなります。

②実験データを読み込むと、下図のように選択した実験データ名・サンプル名等が表示されます。
読み込んだ実験データを削除する場合は「Delete」ボタンをクリックしてください。表示内容に問題がなければ、「Group Edit」か「Next」ボタンをクリックして次へ進みます。
また、この画面での実験データ名の並び順(No の番号)が、メイン画面上でのデータの並び順になります。

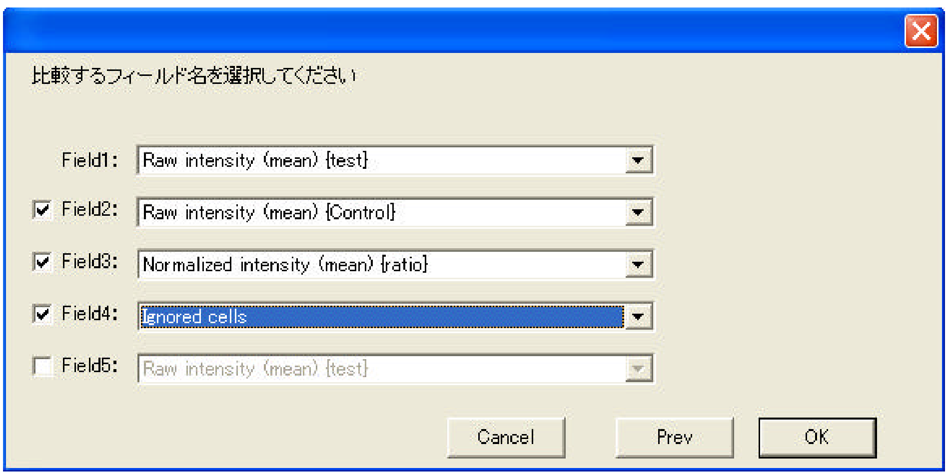


【Group Edit】

複数比較データ解析を行う場合、読み込んだ実験データに対してグループ設定を行うことで、グループ単位でのデータの比較を行うことができます。設定を行うには、「Group Edit」ボタン、またはメイン画面での アイコンをクリックすると下図の画面に移ります。ここで「新規登録」をクリックすると、新しいグループの名前を登録することができます。次にグループ未登録実験を、作成したグループに対して「Add ->」、「<- Remove」ボタンを用いて登録します。グループの登録が終わったら「OK」ボタンをクリックしてください。②の画面に戻ります。

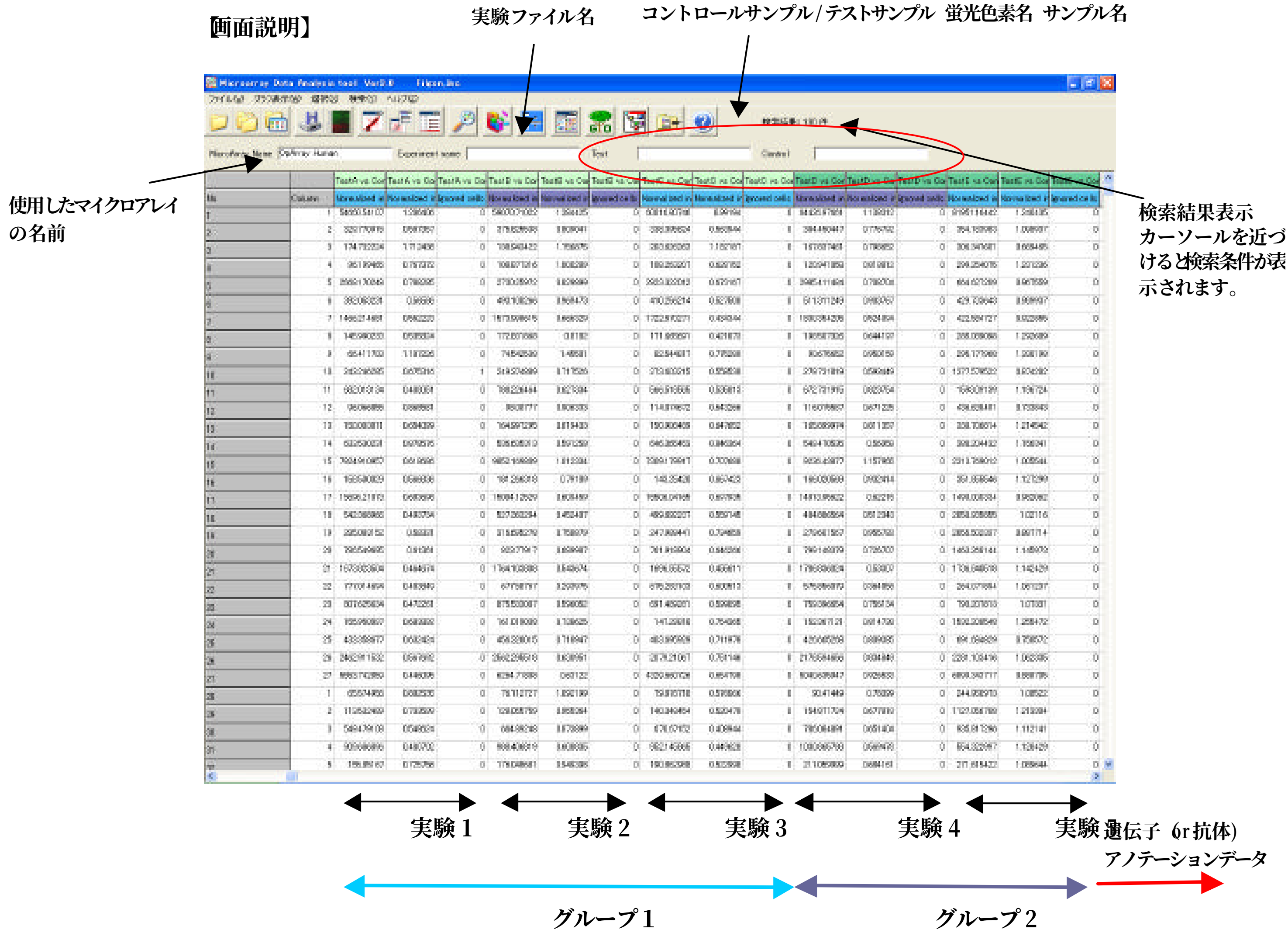


③「Next」ボタンをクリックすると比較するフィールド名を選択する画面が現れます。複数比較の場合、選択したフィールド名のみをメイン画面上に表示させます。検索の際も、ここで選択したフィールドのデータのみ検索対象となります。但し、クラスタリング用出力データ作成の際は「log2ratio」を選択していなくても自動で出力されます。ここでは最大5つのフィールドを選択・表示することができます。「OK」ボタンをクリックするとデータがメイン画面に表示されます。



【複数比較データ解析画面】

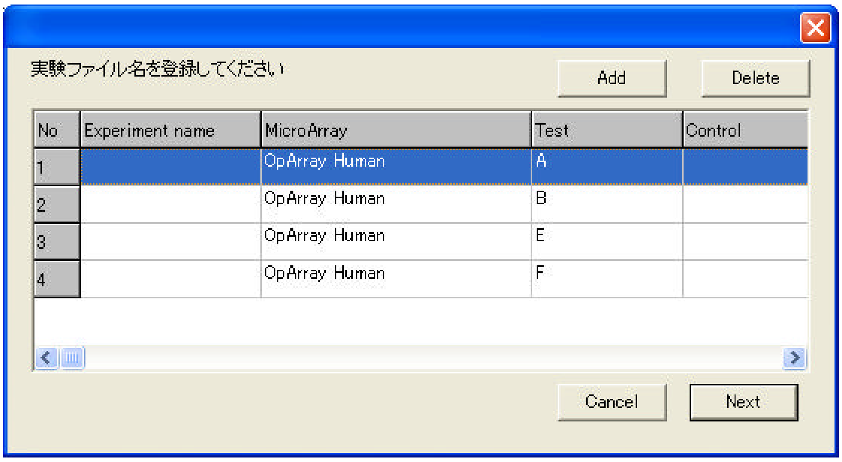
複数比較データの読み込み手順③で選択したフィールドのデータが実験ごとに表示されます。実験ファイル名およびコントロールサンプル/テストサンプルの蛍光色素名とサンプル名は空欄となります。



Signal比較データ解析

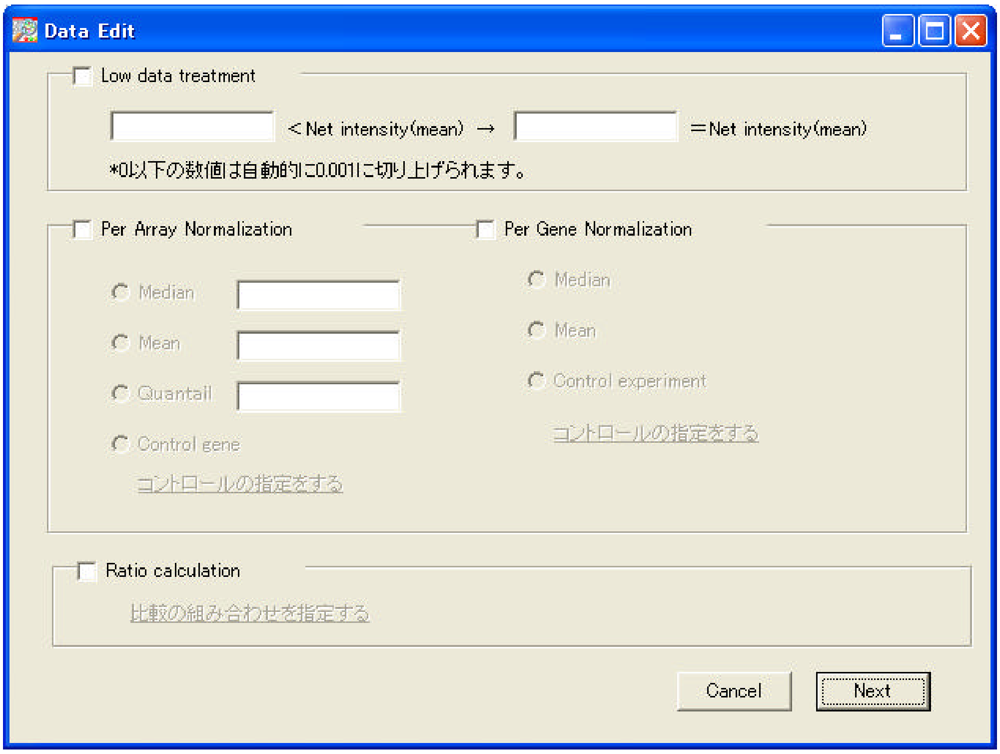
Signal 比較データ解析の場合 ： 2実験以上の単色実験データを対象にデータの加工・検索等の解析を行います。

①アイコンをクリックすると、下記の画面が表示されます。「Add」ボタンをクリックし、対象となる実験データの選択・読み込みを行います。Signal 比較データ解析用の実験データファイルは、2 サンプル分の実験データを含む1比較解析や複数比較解析用のファイルと異なり、1 サンプル分の実験データしかもっていません。そのためファイル名も「Data1 vs Data2」ではなく、「Data1」、「Data2」のように 1 サンプルずつに分かれています。

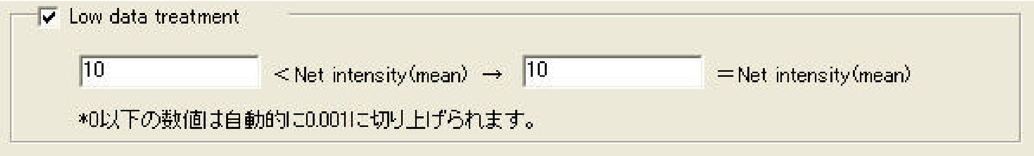


＊複数比較データ解析の場合と違い、「Experiment name」、「Control」、「Control laveling」の項目は空白になっています。

② 「Next」ボタンをクリックすると下の画面に移ります。この画面では、Signal の小さいデータの除去、各種ノーマライズ、サンプル間 Ratio(Fold Change)の計算を行うことができます。これらの機能を使用するには、各項目名の左のチェックボックスにチェックを入れる必要があります。



【Low data treatment】
この項目では Signal の小さいデータ(値がマイナスのデータなど)を一定の値に引き上げることができます。ノーマライズ前のデータである Net intensity (mean) に対して、左側に入力した値未満のデータは右側に入力した値に変換されます。ただし、この機能を使用しなくても、Net intensity (mean) が「0」以下のデータは、自動的に「0.001」にされます。また入力欄に負の値は入力できません。



(例)それぞれの実験データの Net intensity が「10」以下のものを全て「10」に切り上げる場合。

【Per Array Normalization】
複数の単色実験データを比較できるように正規化を行います。Ratio 計算などサンプル間の比較をする場合は、必ずこの処理を行ってください。なお選択できる手法は以下のものがあります。

(1)Median
全ての実験データの中央値を、右側の入力欄で指定した値に揃えます。単色法でのマイクロアレイ実験で、最もよく使用されているノーマライズ法です。

(2)Mean
全ての実験データの平均値を、右側の入力欄で指定した値に揃えます。

(3)Quantile
実験データ内での発現量の順位が同じ場合、全実験データの発現量が同じ値になるように補正します。右側の入力欄に数値を入力すると、入力した値に中央値を揃えますが、値を入力しなかった場合は、「全実験データの中央値の平均値」が中央値となります。実験日が違うデータなどを比較する場合によく使用されます。

(4)Control gene
全実験における、指定遺伝子の発現量を「1」に揃えることによって、指定した遺伝子を基準とする補正を行います。遺伝子の選択には「遺伝子セット」機能を使用するので、予め別の実験データで指定遺伝子の遺伝子セット登録を行ってください。

処理の手順
①ノーマライズや Ratio 計算を行わずに実験データを読み込み、メイン画面に表示させる。
②「Filter Options」の「Keyword」検索機能などを用いてコントロールとしたい遺伝子(GAPDH など)を検索する。
③検索結果であるメイン画面上の遺伝子を「遺伝子セット」として登録する。(P.26 参照)
④一度ソフトウェアを終了させ、再び起動させてから実験データを読み込む。
⑤「Per Array Normalization」で「Control gene」を選択する。
⑥「コントロールの指定をする」をクリックすると、登録されている遺伝子セットの一覧が表示される。
その中から、③で登録したコントロール遺伝子の遺伝子セットを選択する。

この手法では、「全ての実験データにおいて指定した遺伝子の発現量は同じ」ことが前提となっています。そのため、指定遺伝子には「GAPDH」や「β-actin」などのハウスキーピング遺伝子を選択してください。また、選択した遺伝子セット内に複数の遺伝子が登録されている場合は、それらの平均値を使用して補正を行います。

【Per Gene Normalization】
正規化後の発現量に対して、各遺伝子ごとに他の実験データと比較しやすいように標準化を行います。クラスタリングや発現量の統計検定を行うためには、この処理が必要となります。なお、選択できる手法は以下のものがあります。

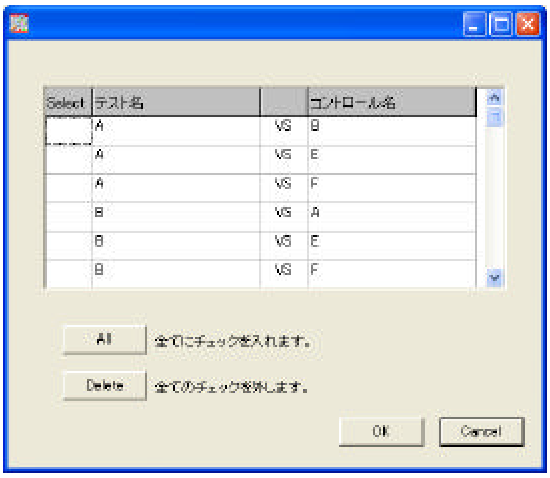
(1)Median
各遺伝子ごとの Per Array Normalized intensity の中央値を使って標準化を行います。

(2)Mean
各遺伝子ごとの Per Array Normalized intensity の平均値を使って標準化を行います。

(3) Control experiment
任意の実験データを指定し、全遺伝子に対して指定した実験データの発現量を基準にして標準化を行います。複数の実験データが指定されたときは、それらの平均値を使用します。

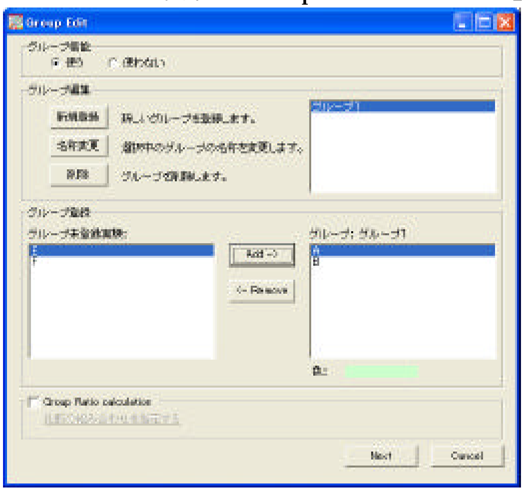
この処理を行うと、各遺伝子の実験データごとの蛍光強度の増減を、「中央値」、「平均値」、「任意の実験データ」を基準(データを「0」とする)にして数値化することができます。このとき出力データはLog2変換されているので、「+」や「−」の符号から、基準に対する増減を直感的に判断することができます。ただし、選択した手法の影響を受けるのはクラスタリング時のヒートマップの色付けに関してだけで、統計検定に関してはどの手法を選択しても結果は同じになります。

【Ratio calculation】

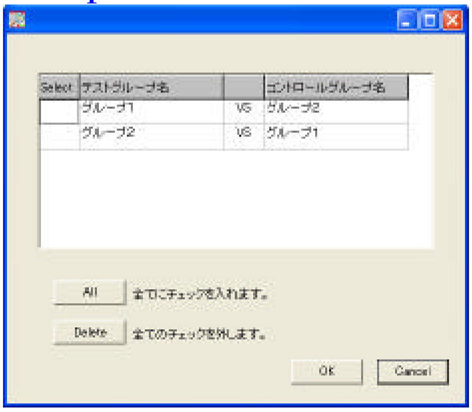


任意の実験データ同士で、各遺伝子ごとの発現比を計算することができます。「比較の組み合わせを指定する」の文字列をクリックすると左のウインドウが開くので、比較したい組み合わせの「Select」列のセルをダブルクリックして選択を行ってください。選択されるとSelect列に「○」が表示されます。ただし、指定した実験データのうちどちらか一方でもIgnored cellsの値が「1」になっている遺伝子は、発現比の計算は行われず、メイン画面上に空白として表示されます。
なお発現比計算では、「テスト名」列の実験データを分子に、「コントロール名」列の実験データを分母にして計算を行います。また計算には「Per Array Normalization」処理後のデータを使用しています。

③「Next」ボタンをクリックすると、グループ設定の画面に移ります。複数比較解析と同様に、各実験データをグループ化することができます。グループ化を行うことで、実験データのグループ別検索、遺伝子発現量のグループ間平均Ratioの計算、さらに統計検定によるグループ間有意差検定を行うことができるようになります。このうちグループ間平均Ratioの計算を行うには、ウインドウ下部の「Group Ratio calculation」にチェックを入れ、設定を行う必要があります。

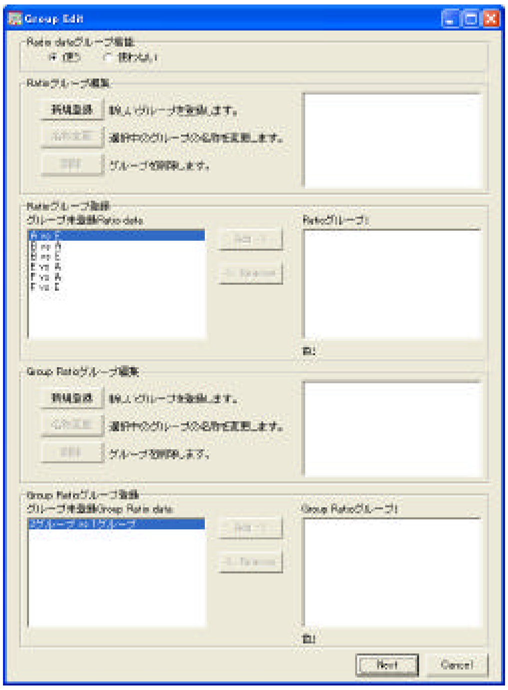


【Group Ratio calculation】



グループ間平均 Ratio 計算を行う場合は、全ての実験データにグループ設定を行った後、「Group Ratio calculation」にチェックを入れてください。「Ratio calculation」と同様、「比較の組み合わせを指定する」の文字列をクリックすると左のウインドウが開くので、比較したい組み合わせの「Select」列のセルをダブルクリックして選択を行ってください。なおグループ間平均 Ratio 計算では、「テストグループ名」列のグループを分子に、「コントロールグループ名」列のグループを分母にして計算を行います。また計算には「Per Array Normalization」処理後のデータの各グループ平均値を使用しています。

④続いてRatioデータのグループ設定を行います。複数のRatioデータをグループ分けすることで、グループ単位での検索を行えるようになります。

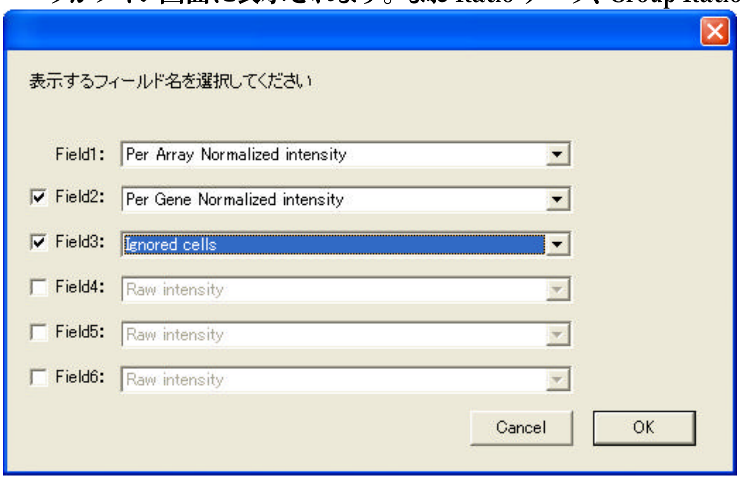


「Signal 比較解析」には 2 種類の Ratio データがあります。1つは「Ratio calculation」で指定した組み合わせの実験データ同士を比較した「Ratio データ」、もう1つは「Group Ratio calculation」で指定した組み合わせの実験データグループ同士を比較した「Group Ratio データ」です。指定した組み合わせが、自動的に画面に表示されるので、実験データのグループ設定と同じ要領で設定を行ってください。

左図のように設定画面は上下に分かれており、上段で Ratio データ、下段で Group Ratio データのグループ分けを行います。

なおRatio データとGroup Ratio データを同じグループに設定することはできません。

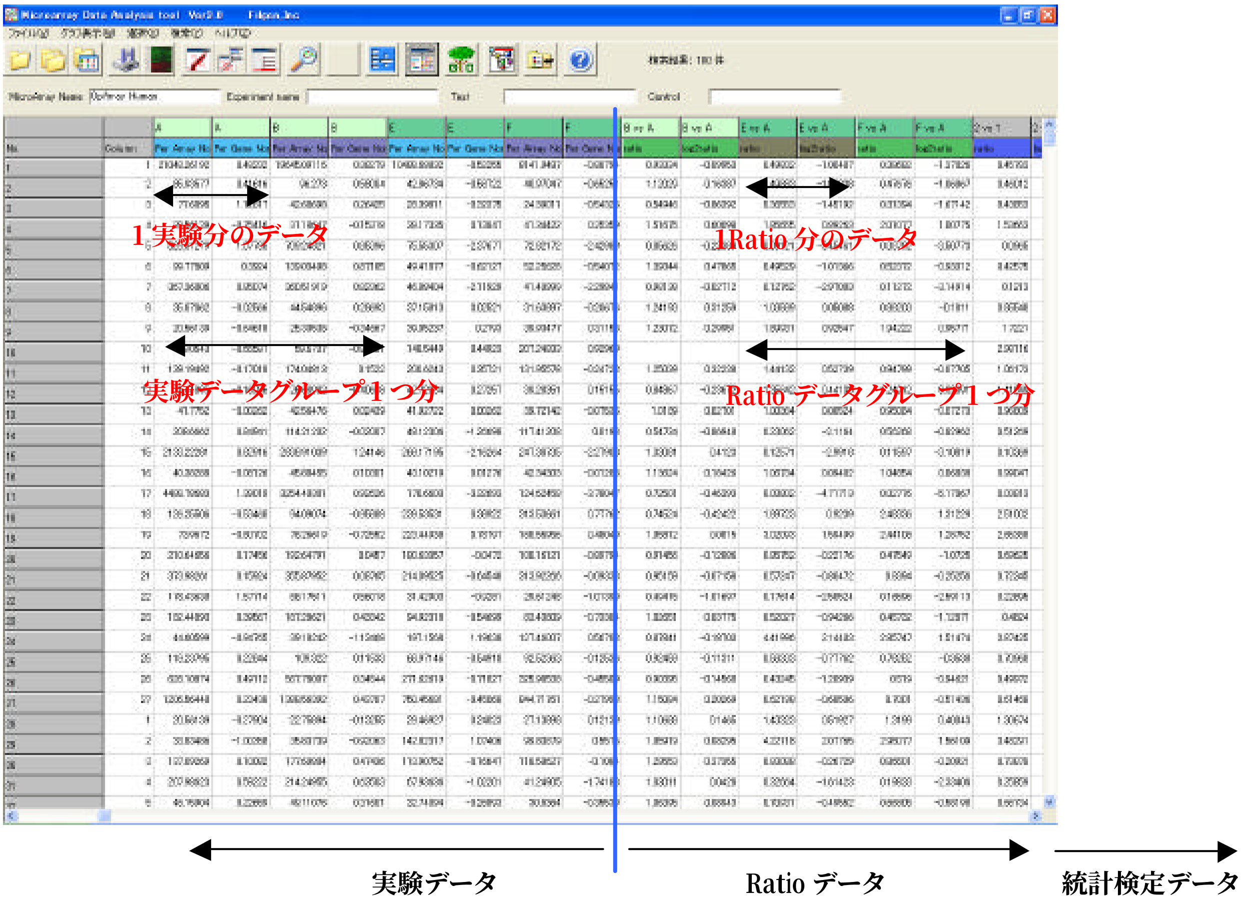
⑤最後にメイン画面に表示するフィールド名を選択する画面が現れます。検索の際も、ここで選択したフィールドのデータのみが検索対象となります。但し、クラスタリング用出力データ作成の際は「Per Gene Normalized intensity」を選択していなくても自動で出力されます。ここでは最大 6 つのフィールドを選択・表示することができます。「OK」ボタンをクリックするとデータがメイン画面に表示されます。なお Ratio データ、Group Ratio データは、指定した組み合わせが全て表示されます。



【Signal比較データ解析画面】

Signal 比較データ解析の読み込み手順⑤で選択したフィールドのデータが実験ごとに表示されます。また②の Ratio calculation、および③の Group Ratio calculation で指定した組み合わせで計算した「Ratio」、「Log2ratio」データも表示されます。実験ファイル名およびコントロールサンプル/テストサンプルの蛍光色素名とサンプル名は空欄となります。

【画面説明】



Signal 比較データ解析での数値データは、大きく分けて 3 種類になります。Net intensity、Per Array Normalized intensity など Signal 強度データから成る「実験データ」、指定した実験データまたはグループデータの組み合わせで計算を行った「Ratio データ」、さらにグループ間の Signal 強度の差を統計的手法で計算した「統計検定データ」の 3 つです。また数値データ以外では、各遺伝子の詳細情報である「アノテーションデータ」も表示されます。

項目名の説明(実験データ)

メイン画面で表示されている実験データの各項目の説明を以下に示します。

【比較、複数比較解析の場合】

項目	説明
No..	行の番号
Select	データの保存や遺伝子(or 抗体)セット登録の対象となる行 ○:対象となる行　空欄:対象とならない行
Block	スポットが搭載されているアレイのBlock番号
Row	スポットが搭載されているアレイのBlockの列番号
Column	スポットが搭載されているアレイのBlockの行番号
Raw intensity (mean) {test } Raw intensity (mean) {control}	Raw intensity (mean) = スポットの intensity(平均値)
Background (mean) {test} Background (mean) {control}	Background (mean) = スポット周囲の intensity(平均値)
Net intensity (mean) {test} Net intensity (mean) {control}	Net intensity (mean) = Raw intensity (mean) - Background (mean)
Net intensity (mean) {sum}	Net intensity (mean) {test}とNet intensity (mean) {control}の合計
Normalized intensity (mean) {test} Normalized intensity (mean) {control}	Normalization 後の Net intensity (mean)
Normalized intensity (mean) {sum}	Normalized intensity (mean) の{test}と{control}の合計
Normalized intensity (mean) {ratio}	Normalized intensity (mean) {test} / Normalized intensity (mean) {control}の値
Normalized intensity (mean) {log2ratio}	Normalized intensity (mean) {ratio}の log2 変換した値
Ignored cells	0 : 解析対象スポット　1 : 解析除外スポット
P-value	グループ間発現量の統計学的検定の有意差を示す確率値
FDR	P-value を Benjamine & Hochberg の手法で補正した False Discovery Rate の値

【Signal比較解析の場合】

項目	説明
No..	行の番号
Select	データの保存や遺伝子(or 抗体)セット登録の対象となる行 ○:対象となる行　空欄:対象とならない行
Block	スポットが搭載されているアレイのBlock番号
Row	スポットが搭載されているアレイのBlockの列番号
Column	スポットが搭載されているアレイのBlockの行番号
Raw intensity	Raw intensity (mean) = スポットの intensity(平均値)
Background	Background (mean) = スポット周囲の intensity(平均値)
Net intensity	Net intensity (mean) = Raw intensity (mean) - Background (mean)
Per Array Normalized intensity	Per Array Normalization 後の intensity
Per Gene Normalized intensity	Per Gene Normalization 後の intensity
ratio	指定の組み合わせで計算した Per Array Normalized intensity の比率
log2ratio	ratio を log2 変換した値
Ignored cells	0 : 解析対象スポット　1 : 解析除外スポット
P-value	発現量の統計学的検定の有意性を示す確率
FDR	P-value を Benjamine & Hochberg の手法で補正した False Discovery Rate の値

＊使用するアレイデータの種類によって上記項目の中で該当するデータがない場合(抗体アレイ、microRNA など平均化データを使用する場合は、Raw データ(Raw intensity、Background)はありません)、または新たな項目が追加されている場合(プロモーターアレイ)があります。

項目名の説明 遺伝子 (or 抗体)アノテーションデータ)

メイン画面で表示されている遺伝子 (or 抗体)アノテーションデータの各項目の説明を以下に示します。
＊下記はＦilgenArrayHuman35kの場合です。使用するマイクロアレイの種類によって付加される遺伝子 (or 抗体)情報は異なります。

項目	説明
oligo_id	Operon Biotechnology 社がプローブ設計をする際に各遺伝子 (or 抗体)を区別するためにつけたオリジナル番号です。 The Human V3 AROS design is based on the Ensembl human 13.31 database. This genelist provides a simplified version of annotation based on the Ensembl human 13.31 database (http://www.ensembl.org).
EnsemblGeneID	Ensembl human 13.31 database の gene_id
Ensembl transcript_id	For common or partial common oligo types, one transcript is chosen randomly as a representative transcript. The id of this representative transcript is the representative_transcript_id. As the individual oligo type represent only one transcript, this transcript is the representative transcript.
oligo_exon_or_transcript	Oligo is fully located in an exon
	Oligo spans more than exon
oligo_type	C : Oligo represents all transcripts of an Ensembl gene
	P : Oligo represents a subset of transcripts of an Ensembl gene
	I : Oligo represents one transcript of an Ensembl gene
	M : Oligo represents multiple Ensembl genes
gene_symbol	Gene symbol approved by HUGO Gene Nomenclature Committee
description_Ensembl	Description based on gene
RefSeq	RefSeq accession from NCBI Reference Sequence collection which is mapped to the human genome by Ensembl. For further information on RefSeqs, please refer to http://www.ncbi.nlm.nih.gov/RefSeq/.
description_RefSeq	Description based on gene
GB_accession	Accession from EMBL and GenBank
LocusLink	LocusLinkID
homolog	homolog_mouse EnsemblGeneID
	homolog_rat EnsemblGeneID
Gene Ontology	GO_biological_process
	GO_molecular_function
	GO_cell_location
Chromosome	ChromosomeName
	Start Position(bp)
	End Position(bp)
	Middle Position(bp)
	Mb
	Strand
＊ oligo_id 以外すべて空欄のプローブがあります。これは、設計に使用したデータベース (Ensembl)の更新により、設計時には遺伝子あるいは遺伝子候補と思われたものが、変更されたためです。	

ハイパーリンク機能



遺伝子（or 抗体）アノテーションデータの中で青文字表記のデータをダブルクリックすると、対応する公共データベース等にハイパーリンクされます。但し、1つのセルの中に複数のデータがある場合、あるいは「―」、「Blank」の場合にはハイパーリンク機能は対応していません。

Select機能について

Filter Option での検索や検索データの保存などは、すべて Select ボックスにチェックが入っているデータを対象とします。Select ボックスのチェックの入れ方は、Select ボックスをダブルクリックする他に、右クリックすることで下記の選択メニューから設定することができます。

指定の行を選択(W)　：選択しているセルが含まれる行を選択します。「Select」の項に「〇」がつきます。

指定の行を未選択(X)　：選択しているセルが含まれる行を未選択にします。「Select」の項の「〇」が消えます。

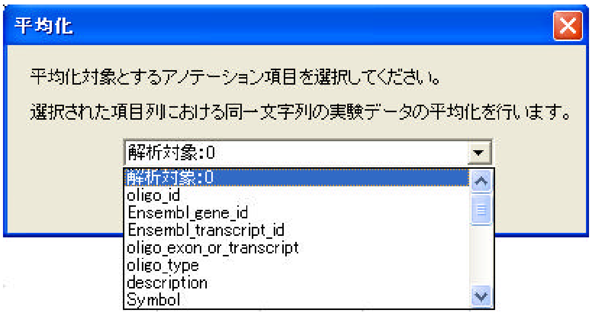
全て選択(Y)　：表示されているすべての行を選択します。すべての「Select」の項に「〇」がつきます。

全て未選択(Z)　：表示されているすべての行を未選択にします。すべての「Select」の項の「〇」が消えます。

平均化機能について

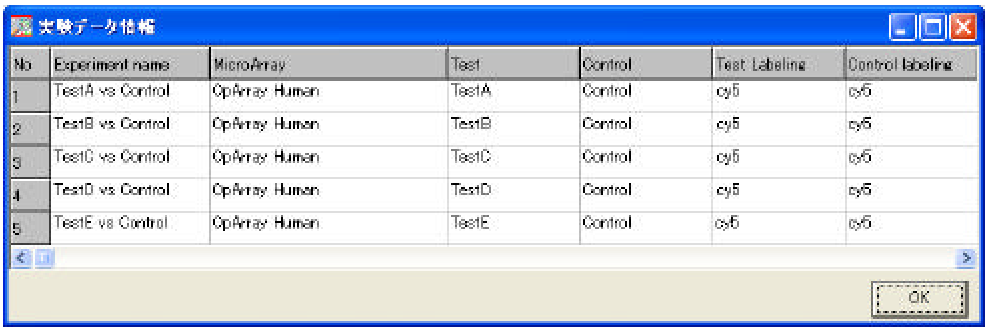
メイン画面のデータは Oligo ID などの特定のアノテーション項目を基準にして、同一の ID のデータをまとめて平均化することができます。このとき Ignored cells の項目が 1 になっているデータは、平均化処理から除外して計算が行われます。また、アノテーション項目が空白になっている場合は平均化の対象外になります。

メイン画面の平均化アイコンをクリックすると下図の画面が表示されますので、ここで平均化の基準にしたいアノテーション項目を選択してください。平均化を実行すると、指定したアノテーション項目において同一の文字列の実験データ（Net intensity、Normalized intensity など）が平均化され、同一のアノテーションをもつ実験データの中で一番 No. が小さい実験データ行に平均化データが反映されます。この時、他の実験データ行はメイン画面から削除されます。なお一度平均化処理を行うと、平均化前の状態に戻せなくなります。元に戻したい場合は、実験データを再び読み込み直す必要があります。



実験データの表示について

メイン画面左上のアイコンをクリックすると、現在読み込まれている実験データのサンプル情報が表示されます。



実験データ情報

No	Experiment name	MicroArray	Test	Control	Test Labeling	Control labeling
1	TestA vs Control	OpArray Human	TestA	Control	cy5	cy5
2	TestB vs Control	OpArray Human	TestB	Control	cy5	cy5
3	TestC vs Control	OpArray Human	TestC	Control	cy5	cy5
4	TestD vs Control	OpArray Human	TestD	Control	cy5	cy5
5	TestE vs Control	OpArray Human	TestE	Control	cy5	cy5

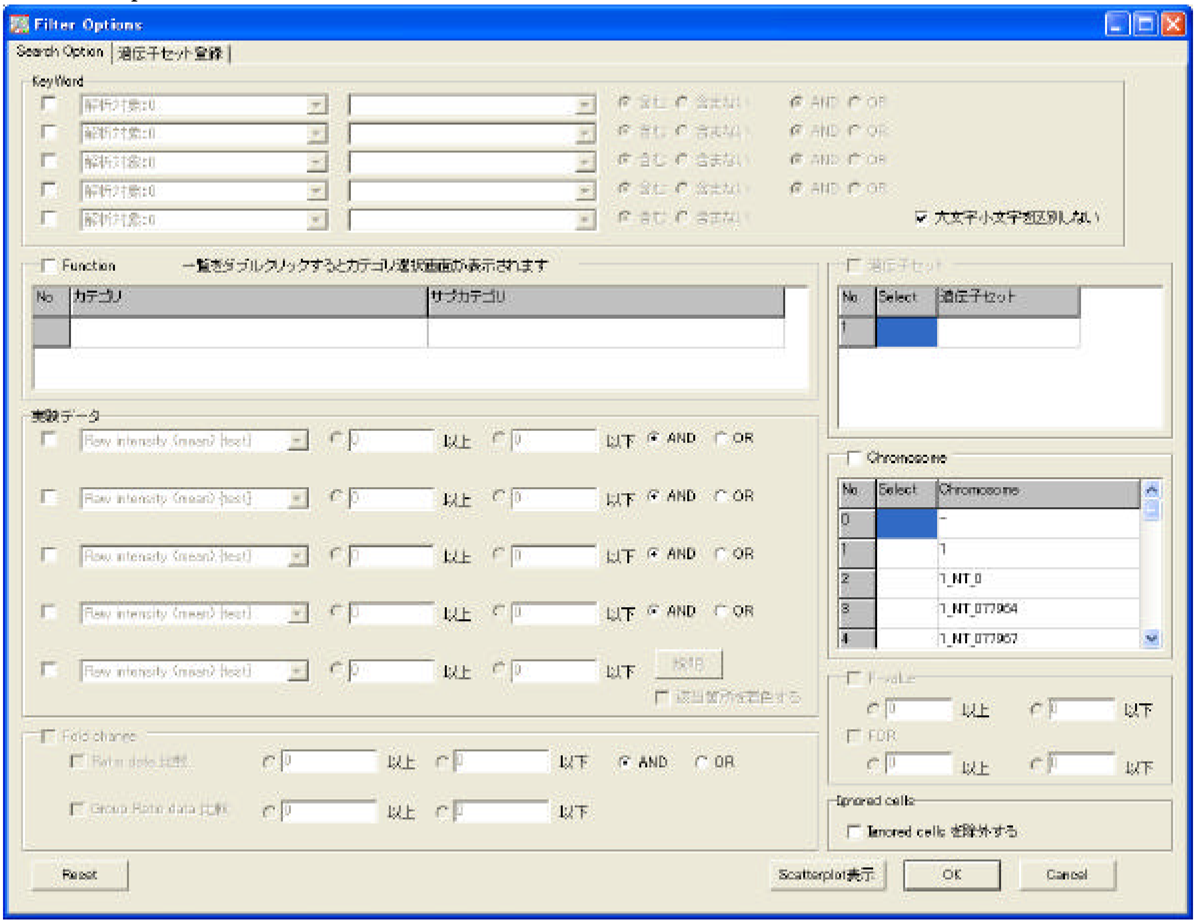
OK

4. ソフトウェアの操作　検索（Filter Options）

Filter Options 画面

Filter Options画面では1比較データ、複数比較データ、Signal 比較データを対象に、各種検索条件を設定し、データの抽出を行います（Search Option）。また、選択した遺伝子（or 抗体）群を遺伝子（or 抗体）セットとして登録することができます(遺伝子（or 抗体）セット登録)。

Filter options アイコンをクリックすると下記のような画面が表示されます。



【Search Option】 Search Option のタブでは各種（Keyword、実験データ、P-value など）検索条件を設定し、データの抽出を行います。各検索は一度に複数行うことができます。

【遺伝子セット登録】 検索対象とする遺伝子（or 抗体）セットを選択することができます。遺伝子（or 抗体）セットの選択を行う場合は、最初に遺伝子セットのチェックボックスをクリックし、登録されている遺伝子セットから対象とする遺伝子（or 抗体）セットを選択します。詳細は P.27【遺伝子セット登録について】をご参照ください。

【Search Optionの検索項目】

Keyword検索

:チェックボックスをクリックし、対象となるアノテーション項目をプルダウンメニューから選択します。

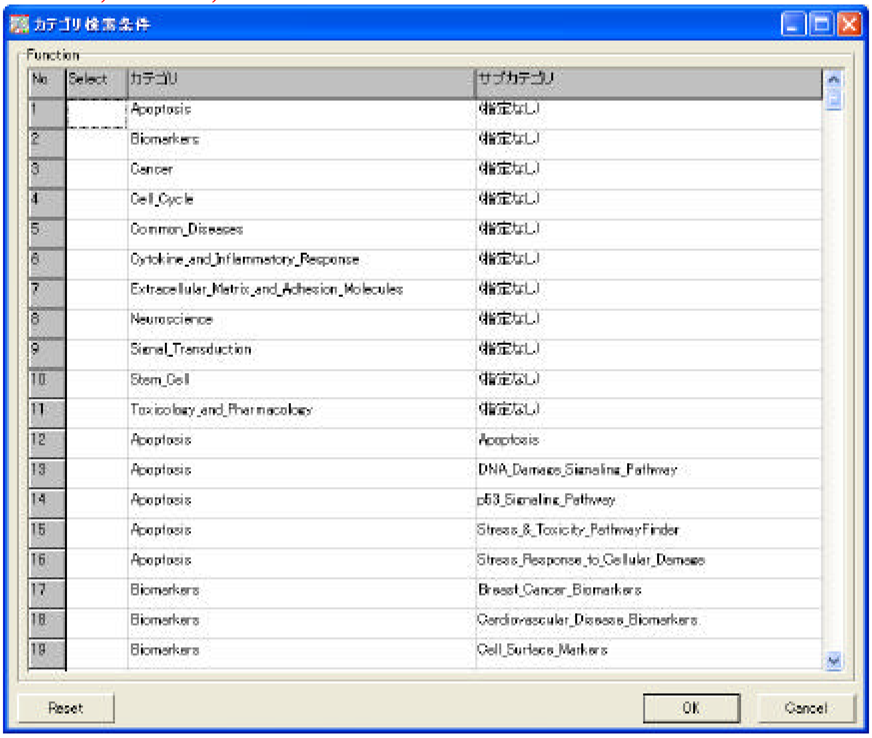
KeyWord を入力し、検索を行います。最大 5 つの KeyWord 検索ができます。また、KeyWord 一つ一つに対する AND 検索、OR 検索の指定や、入力した KeyWord に対して、大文字小文字の区別をする、しないの選択ができます。

遺伝子の検索を行うときには、まず最初に「解析対象 :0」の項目でブランクなどを除いた解析対象遺伝子のみを選択する必要があります。このとき選択できる各数字(0、2、3)は、それぞれ 0 :解析対象スポット、2 :コントロールまたはブランクスポット、3 :データベースの更新によりアノテーション情報が削除された遺伝子のスポットを表しています。よって解析時には、最初に「解析対象 :0」で「0」を選択し、解析対象の遺伝子のみを検索します。アレイの種類により、「解析対象 :0」の項目がない場合があります。

Function検索（Human, Mouse, Ratのみ）

:カテゴリー別に分類された遺伝子（or 抗体）群の検索を行います。
カテゴリー検索を行う場合は、最初にFunction のチェックボックスをクリックしてください。
分類には大まかなカテゴリー分類と、各カテゴリーをさらに分類したサブカテゴリーがあります。つまり大カテゴリーにおける該当遺伝子数は、その全てのサブカテゴリーの該当遺伝子数の合計と同じになります。検索時にはカテゴリー項目のSelect ボックスをダブルクリックしてください。また、カテゴリーは複数選択を行うことができます。

Human, Mouse, Ratのみのアレイに対応する機能です。



Choromosome検索

染色体の番号別の検索を行います。
染色体検索を行う場合は、最初にChromosome のチェックボックスをクリックします。
スクロールで目的の染色体番号を見つけ出し、対象となる染色体番号の Select ボックスをダブルクリックします。
また、染色体番号の複数選択も行うことができます。

実験データ検索

選択した実験数値データの項目を対象に、条件検索を行います。
最大５つの設定条件を設け、それぞれの条件に対してAND、OR検索を指定することができます。
上に入力されている条件から順に検索していくので、特にOR検索を行う場合は順番に気をつけてください。入力する位置によって検索結果が変わってくる場合があります。

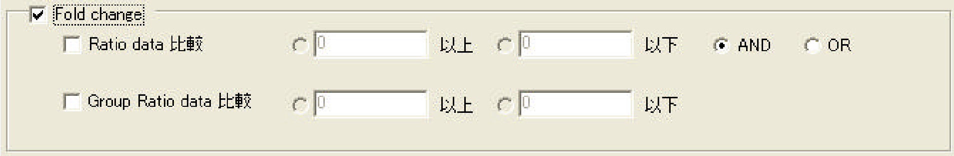
例）Normalized intensity が 100以上で、Ratioが 2倍以上、または0.5倍以下を設定する場合。



*ここでRatioが 2倍以上、または0.5倍以下の項目をintensity が 100以上の後にすると、検索結果が変わってしまいます。

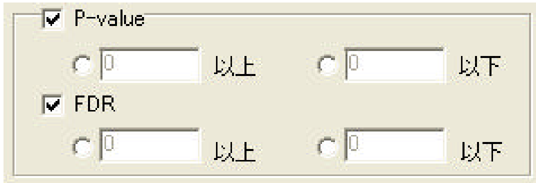
Fold change検索

Signal 比較データ解析のみで使用できる機能です。指定の組み合わせで計算した Ratio や Group Ratio data に対して検索を行うことができます。複数データに対する検索では、実験データ検索と同じようにグループ別検索を行うことができます。
検索を行うには、まず Fold change のチェックボックスにチェックを入れます。次に実験間個別の Ratio に対して検索を行う場合は「Ratio data 比較」に、また実験データグループ内の平均 Signal 値を用いて計算した Group Ratio に対して行う場合は「Group Ratio data 比較」にチェックを入れ、数値を入力してください。また、必要に応じて複数比較の条件を指定してください。



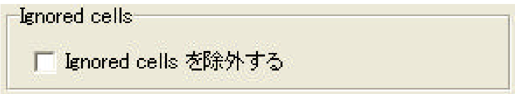
P-value、FDR検索

：統計検定」機能で計算した各遺伝子に対する P-value、または FDR の検索を行います。ここで P-value または FDR のチェックボックスをクリックし、検索条件を指定することで、統計的有意性をもつ遺伝子のみを抽出することができます。



Ignored cells除外設定

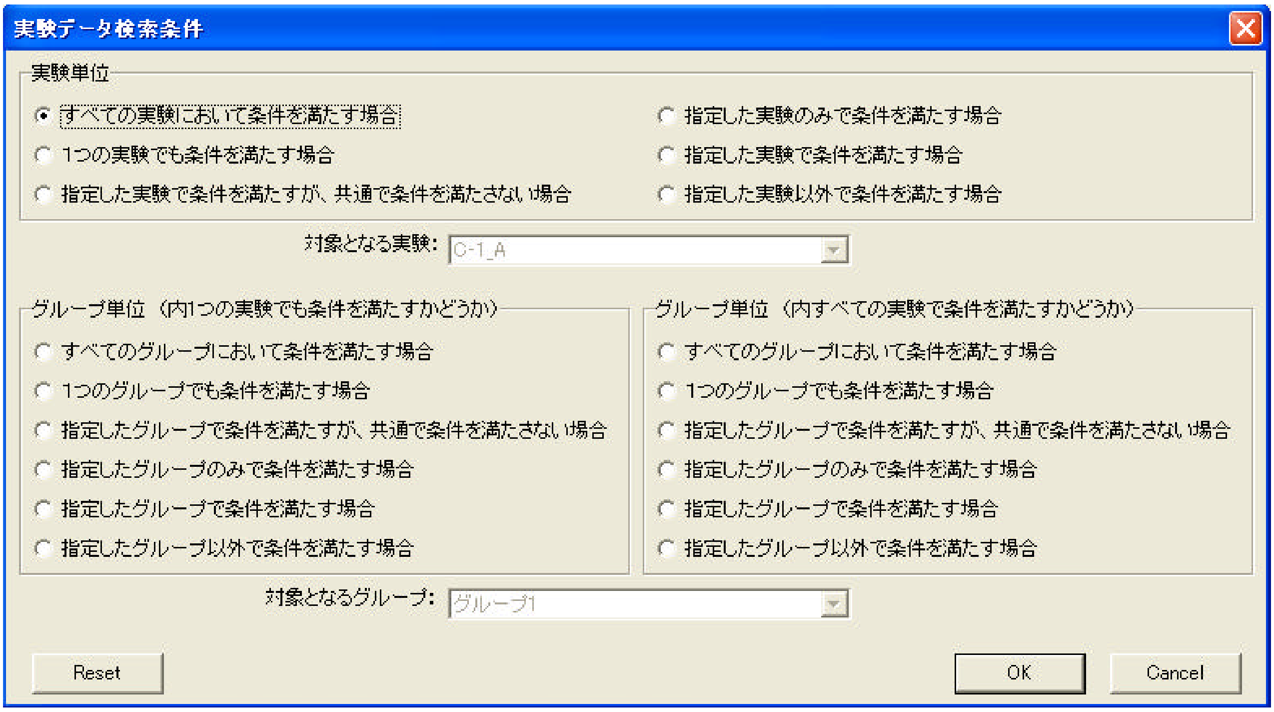
：解析除外スポットを指定している Ignored cells の項目が「1」になっている遺伝子を、検索から除外することができます。ただし複数または Signal 比較解析の場合、1 実験でも Ignored cells が「1」になっているものがあると、他の実験データが「0」でも除外されてしまいますのでご注意してください。



複数データに対する検索条件

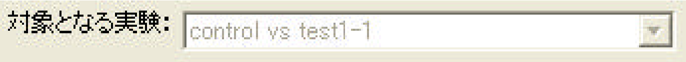
複数または Signal 比較データ解析の場合、2 つ以上の実験データや Ratio データが存在するため、指定した検索をどのデータあるいはグループを対象として実行するかを指定することができます。

検索の種類は下図の 18 種類となります。

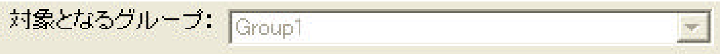


＊Signal 比較データ解析の Fold change 検索でも、上記と同様の検索設定を行うことができます。その場合でのグループとは、Ratio データグループとなります。

・実験データ、グループの指定



複数データの検索で、検索対象とする実験データの選択を行います。プルダウンメニューから指定する実験データを選択してください。



複数データの検索で、検索対象とするグループの選択を行います。プルダウンメニューから指定するグループを選択してください。

・データ画面の着色について



「該当箇所を着色する」にチェックを入れると、検索後のメイン画面において、指定したする検索条件を満たす実験データのセルを赤色に着色することができます。

＊「該当箇所を着色する」を選択した場合、検索結果が表示されるまでの時間が長くなることがあります。



Reset ボタン　：選択されている検索条件をすべて解除し、初期状態に戻します。



OKボタン　：選択されている検索条件を実行します。



Cancelボタン　：選択されている検索条件を実行せず、Filter option 画面を閉じます。

・検索条件

複数比較実験を行う時の、検索条件には以下のような項目があります。

【実験単位】

すべての実験において条件を満たす場合

- すべての実験において指定した検索条件を満たすデータを抽出します。
  例えば、Ratioが 2 倍以上（Up変動）という条件で検索した場合、実験A、実験B、実験Cすべての実験でRatioが 2 倍以上のデータを抽出します。
  コントロールが共通であるような関連実験データでは、この検索により共通に変動する遺伝子（or 抗体)を簡単にピックアップすることができます。一方で、変動しているが共通に変動していない遺伝子（or 抗体)は、その実験特異的に変動している遺伝子（or 抗体)となります。これらの遺伝子（or 抗体)群を抽出する場合は「指定した実験で条件を満たすが、共通で条件を満たさない場合」あるいは、「指定した実験のみで条件を満たす場合」を選択してください。

1つの実験でも条件を満たす場合

- 1 つの実験でも指定した検索条件を満たすデータを抽出します。
  例えば、Ratioが 2 倍以上（Up変動）という条件で検索した場合、実験A、実験B、実験Cのうち 1 つでもRatioが 2 倍以上のデータがあれば、そのデータを抽出します。

指定した実験で条件を満たすが、共通で条件を満たさない場合

- 指定した実験データにおいて指定した検索条件を満たすが、すべての実験で共通で条件を満たしているデータがある場合はそれらを省いてデータを抽出します。
  例えば、実験Aを指定し、Ratioが 2 倍以上（Up変動）という条件で検索した場合、実験Aで指定した条件を満すデータを抽出しますが、その中で実験B、実験CでもRatioが 2 倍以上のデータがあれば、これらのデータを除いたデータが抽出されます。

指定した実験のみで条件を満たす場合

- 指定した実験のみで指定した検索条件を満たしているデータを抽出します。
  例えば、実験Aを指定し、Ratioが 2 倍以上（Up変動）という条件で検索した場合、実験Aで指定した条件を満すデータを抽出しますが、実験B、実験Cのどちらか一方でもRatioが 2 倍以上のデータがあれば、これらのデータを除いたデータが抽出されます。

指定した実験以外で条件を満たす場合

・ 指定した実験以外で指定した検索条件を満たしているデータを抽出します。
例えば、実験Aを指定し、Ratioが2倍以上（Up変動）という条件で検索した場合、実験Aでは指定した条件を満すデータを抽出しませんが、実験B、実験Cのどちらか一方でもRatioが2倍以上のデータがあれば、そのデータを抽出します。

指定した実験で条件を満たす場合

・ 指定した実験で指定した検索条件を満たしているデータを抽出します。
例えば、実験Aを指定し、Ratioが2倍以上（Up変動）という条件で検索した場合、実験A内から指定した条件を満すデータを抽出します。

【グループ単位（内１つの実験でも条件を満たすかどうか）】

すべてのグループにおいて条件を満たす場合

・ すべてのグループにおいて指定した検索条件を満たすデータを抽出します。
この場合、グループ内で１つでも検索条件を満たすデータがあれば抽出します。
例えば、Ratioが2倍以上（Up変動）という条件で検索した場合、グループA、グループB、グループCすべてのグループでRatioが2倍以上になっているデータを抽出します。

1つのグループでも条件を満たす場合

・ １つのグループでも指定した検索条件を満たすデータを抽出します。
この場合、グループ内で１つでも検索条件を満たすデータがあれば抽出します。
例えば、Ratioが2倍以上という条件で検索した場合、グループA、グループB、グループCのうち１つでもRatioが2倍以上のデータがあれば、そのデータを抽出します。

指定したグループで条件を満たすが、共通で条件を満たさない場合

・ 指定したグループにおいて指定した検索条件を満たすが、すべてのグループで共通で条件を満たしているデータがある場合はそれらを省いてデータを抽出します。
この場合、指定したグループ内で１つでも検索条件を満たすデータがあれば抽出します。
例えば、グループAを指定し、Ratioが2倍以上という条件で検索した場合、グループAで指定した条件を満たすデータを抽出しますが、その中でグループB、グループCでもRatioが2倍以上のデータがあれば、これらのデータを除いたデータが抽出されます。

指定したグループのみで条件を満たす場合

・ 指定したグループのみで指定した検索条件を満たしているデータを抽出します。
この場合、グループ内で１つでも検索条件を満たすデータがあれば抽出します。
例えば、グループAを指定し、Ratioが2倍以上（Up変動）という条件で検索した場合、グループA内で１つでもRatioが2倍以上になっており、なおかつ他のグループでは2倍以上にはなっているものがないデータを抽出します。

指定したグループで条件を満たす場合

・ 指定したグループで指定した検索条件を満たしているデータを抽出します。
この場合、グループ内で１つでも検索条件を満たすデータがあれば抽出します。
例えば、グループAを指定し、Ratioが2倍以上（Up変動）という条件で検索した場合、グループA内で１つでもRatioが2倍以上になっていればデータを抽出します。この場合、他のグループは考慮に入れません。

指定したグループ以外で条件を満たす場合

・ 指定したグループ以外で指定した検索条件を満たしているデータを抽出します。
この場合、指定したグループでは条件を満たすデータがなく、他のグループ内で１つでも検索条件を満たすデータがあれば抽出します。
例えば、グループAを指定し、Ratioが2倍以上（Up変動）という条件で検索した場合、グループAでは2倍以上にはなっておらず、他のグループのうちどこか1つでも2倍以上になっていればデータを抽出します。

【グループ単位（内すべての実験で条件を満たすかどうか）】

すべてのグループにおいて条件を満たす場合

- すべてのグループにおいて指定した検索条件を満たすデータを抽出します。
  この場合、グループ内すべての実験で検索条件を満たすデータがあれば抽出します。
  例えば、Ratio が 2 倍以上（Up 変動）という条件で検索した場合、グループ A、グループ B、グループ C すべてのグループで Ratio が 2 倍以上になっているデータを抽出します。

1 つのグループでも条件を満たす場合

- 1 つのグループでも指定した検索条件を満たすデータを抽出します。
  この場合、グループ内すべての実験で検索条件を満たすデータがあれば抽出します。
  例えば、Ratio が 2 倍以上という条件で検索した場合、グループ A、グループ B、グループ C のうち 1 つでも Ratio が 2 倍以上のデータがあれば、そのデータを抽出します。

指定したグループで条件を満たすが、共通で条件を満たさない場合

- 指定したグループにおいて指定した検索条件を満たすが、すべてのグループで共通で条件を満たしているデータがある場合はそれらを省いてデータを抽出します。
  この場合、グループ内すべての実験で検索条件を満たすデータがあれば抽出します。
  例えば、グループ A を指定し、Ratio が 2 倍以上という条件で検索した場合、グループ A で指定した条件を満たすデータを抽出しますが、その中でグループ B、グループ C でも Ratio が 2 倍以上のデータがあれば、これらのデータを除いたデータが抽出されます。

指定したグループのみで条件を満たす場合

- 指定したグループのみで指定した検索条件を満たしているデータを抽出します。
  この場合、グループ内すべての実験で検索条件を満たすデータがあれば抽出します。
  例えば、グループ A を指定し、Ratio が 2 倍以上（Up 変動）という条件で検索した場合、グループ A 内すべてで Ratio が 2 倍以上になっており、なおかつ他のグループでは 2 倍以上になっているものがないデータを抽出します。

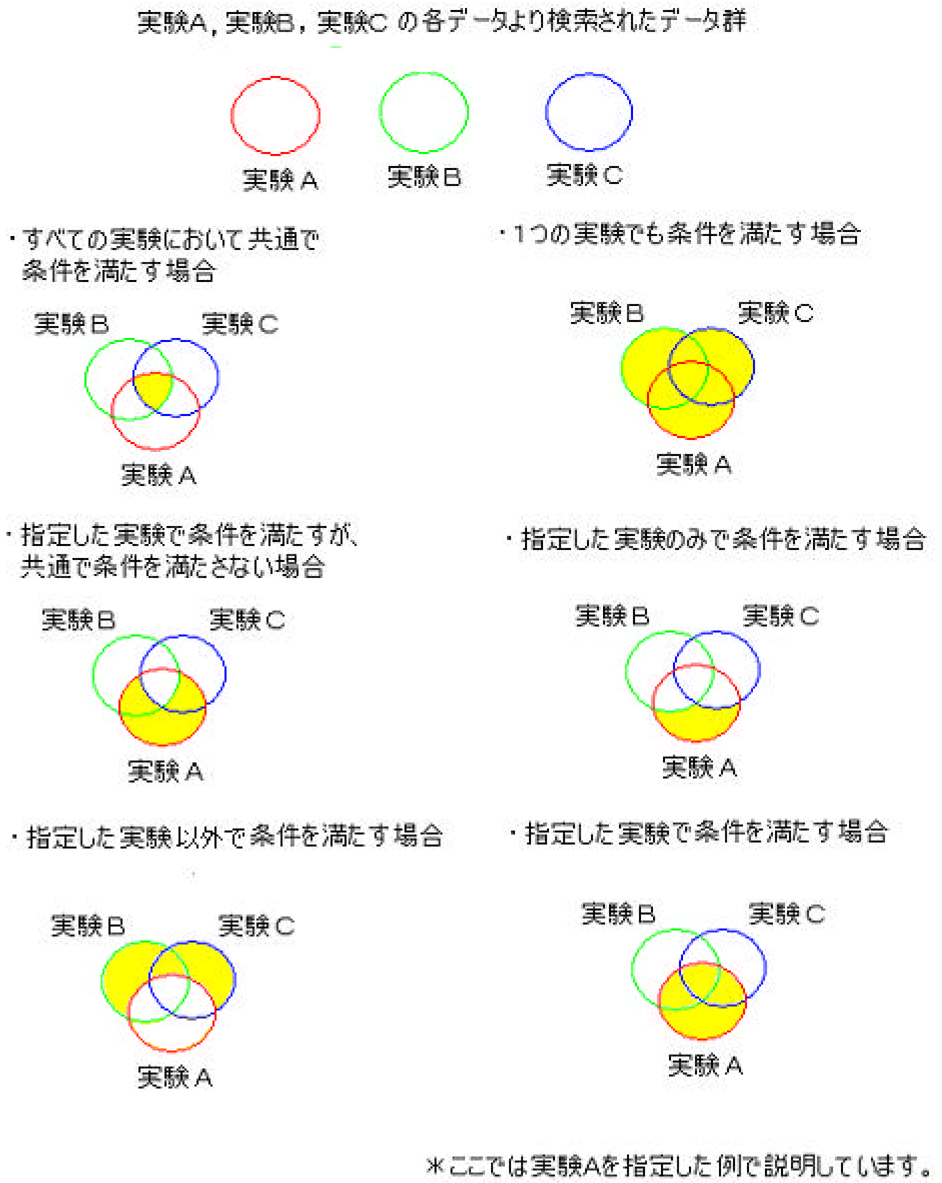
指定したグループで条件を満たす場合

- 指定したグループで指定した検索条件を満たしているデータを抽出します。
  この場合、グループ内すべての実験で検索条件を満たすデータがあれば抽出します。
  例えば、グループ A を指定し、Ratio が 2 倍以上（Up 変動）という条件で検索した場合、グループ A 内すべてで Ratio が 2 倍以上になっていればデータを抽出します。この場合、他のグループは考慮に入れません。

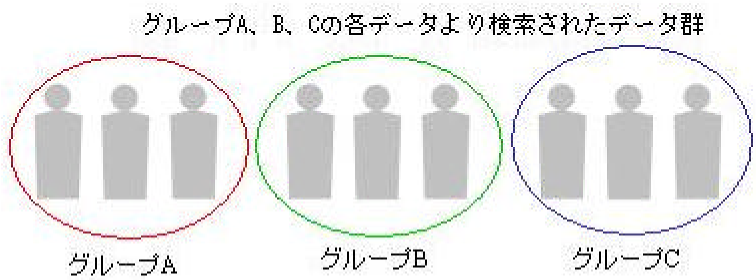
指定したグループ以外で条件を満たす場合

- 指定したグループ以外で指定した検索条件を満たしているデータを抽出します。
  この場合、指定したグループでは条件を満たすデータがなく、他のグループ内すべての実験で条件を満たすデータがあれば抽出します。
  例えば、グループ A を指定し、Ratio が 2 倍以上（Up 変動）という条件で検索した場合、グループ A では 2 倍以上にはなっておらず、他のグループ内すべての実験で 2 倍以上になっていればデータを抽出します。

【複数比較データ解析における検索の概要イメージ図】



【複数グループ比較データ解析における検索の概要イメージ図】



内1つの実験でも条件を満たすかどうか

・すべてのグループにおいて条件を満たす場合



・指定したグループのみで条件を満たす場合



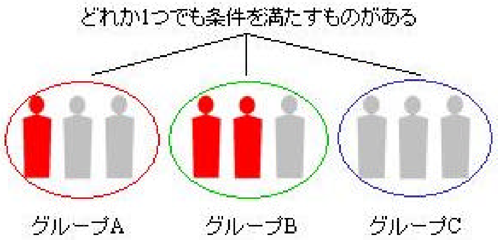
・指定したグループで条件を満たす場合



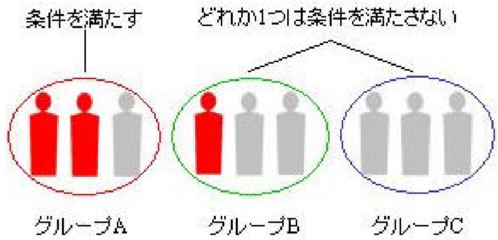
・指定したグループ以外で条件を満たす場合



・1つのグループでも条件を満たす場合



・指定したグループで条件を満たすが、共通で条件を満たさない場合



内すべての実験で条件を満たすかどうか

・すべてのグループにおいて条件を満たす場合



・指定したグループのみで条件を満たす場合



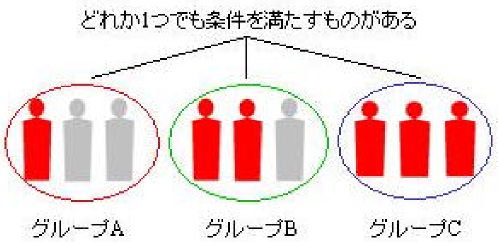
・指定したグループで条件を満たす場合



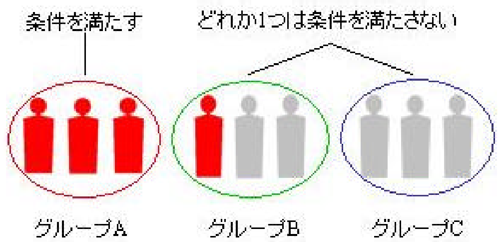
・指定したグループ以外で条件を満たす場合



・1つのグループでも条件を満たす場合



・指定したグループで条件を満たすが、共通で条件を満たさない場合



*ここではグループＡを指定した例で説明しています。

実験データ設定例

実験データ検索では発現変動の大きい遺伝子（or 抗体）の抽出や信頼性の低いデータの除去（CutOff）を簡単に行うことができます。
下記に設定の 1例を記します。アレイの種類や実験内容や変動遺伝子（or 抗体）の数によって、設定を変更してください。

【解析対象遺伝子の設定】： 遺伝子の検索を行うときには、まず最初に KeyWord 検索でコントロール遺伝子などを除いた解析対象遺伝子のみを選択する必要があります。「解析対象 :0」の項目を選択し、「0」を選択してください。
設定方法は P.16 KeyWord 検索の項をご参照ください。

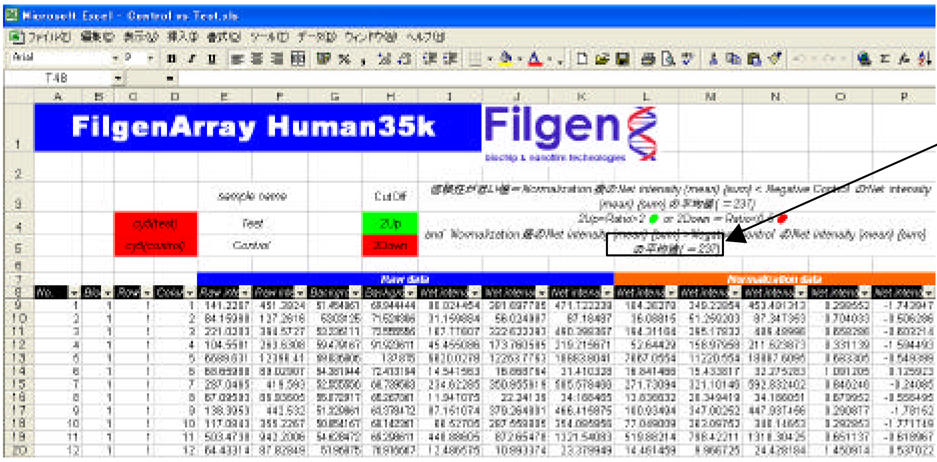
【発現変動遺伝子の設定】： 遺伝子発現解析におけるマイクロアレイの解析では通常コントロールとテストサンプルの intensity の差が 2 倍以上（Up 変動）、0.5 倍以下（Down 変動）の場合において、有意な変動であるとみなします。実験手法やアレイの種類により、変動 Ratio の設定は異なります。
＊変動遺伝子（or 抗体）が多い場合は、必要に応じて、設定値を上げてください。

・1 比較、複数比較解析の場合：有意な Up 変動を抽出する場合は Normalized intensity(mean){ ratio }で「2 以上」、有意な Down 変動を抽出する場合は「0.5 以下」となるように設定してください。
＊Normalized intensity(mean){ log2ratio }で「1 以上」、「-1 以下」でも同様の検索ができます。

・Signal 比較解析の場合： Fold change 検索を用いて、指定の組み合わせで計算した Ratio に対して、Up 変動では「2 以上」、Down 変動を抽出する場合は「0.5 以下」となるように設定してください。なお Fold change 検索では、log2ratio データに対する検索はできません。

【CutOff の設定】： バックグラウンドレベルの信頼性の低いデータを除去することは、マイクロアレイの解析において重要なポイントになります。下記は弊社で行っています一般的な CutOff の設定例です。
*変動遺伝子（or 抗体）が多い場合は、必要に応じて、CutOff の設定値を上げてください。

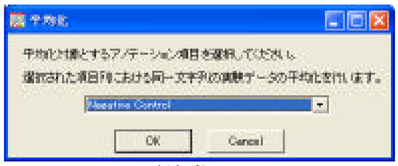
・1 比較解析の場合： 1比較データ解析の場合、テストサンプルとコントロールサンプルそれぞれのネガティブコントロールプローブの蛍光強度の平均値を目安に CutOff 値を設定します。ネガティブコントロールプローブの平均値は、納品用 Excel データ 3行目の CutOff の項目をご覧ください。
設定方法はプルダウンメニューから Normalized intensity(mean){sum}を選択し、「以上」側の数値入力欄に、対応する実験データのネガティブコントロールの CutOff 値を入力します。



・複数比較解析の場合： 実験データを複数読み込む複数比較解析の場合、複数の Excel データシートを確認しなければいけません。例えば 2実験分のデータを読み込む場合、その組み合わせの Excel データを開き、ネガティブコントロールの CutOff 値を確認します。それぞれの CutOff 値が 100、もう一つが 200 となっていた場合、2実験のうち値の小さい方、つまり 100 を CutOff 値として使用します。
設定方法はプルダウンメニューから Normalized intensity(mean){sum}を選択し、「以上」側の数値入力欄に、対応する実験データのネガティブコントロールの CutOff 値を入力します。そして複数データの検索条件で「全ての実験に対して条件を満たす場合」を選択してください。

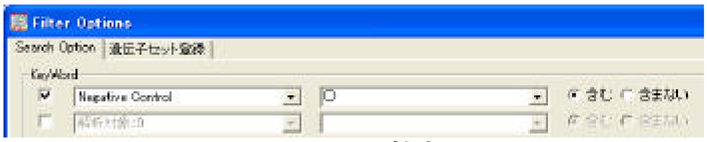
・Signal 比較解析の場合：Signal 比較解析では、テストサンプルとコントロールサンプルの蛍光強度の合計値ではなく 1 サンプル(実験)ごとに Cutoff を行います。ただし Signal 比較解析では、ネガティブコントロールの平均値を自分で計算しなければいけません。方法は、まず平均化アイコン をクリックし、項目の中から「Negative Control」を選択し、平均化処理を実行します。その後、「Filter Options」の Keyword 検索で、「Negative Control」に対して「○」を抽出するように設定を行います。すると各実験データに対して、ネ

ガティブコントロールの平均値が出力されるので、そのうちの Per Array Normalized intensity の最小値をCutOff 値にします。
設定方法は、実験データ項目のプルダウンメニューからPer Array Normalized intensity を選択し、「以上」側の数値入力欄に、CutOff 値を入力します。そして複数データの検索条件で「全ての実験に対して条件を満たす場合」を選択してください。



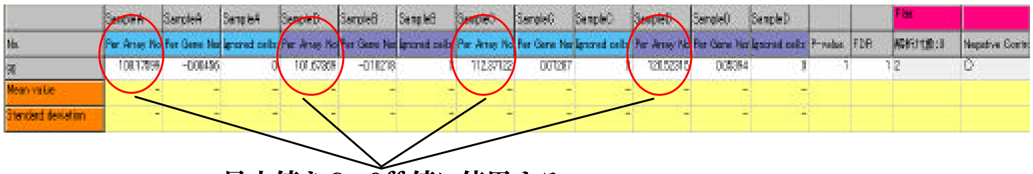
平均化

↓



データの抽出

↓



最小値をCutOff 値に使用する

【解析対象外スポットの除去】：ignore cells の項目では解析対象データには「0」、解析対象外のデータには、「1」と入力してあります。Ignored cells 除外設定の項目で、「Ignored cells を除外する」にチェックを入れてください。ただし複数または Signal 比較解析の場合、1 実験でもIgnored cells が「1」になっているものがあると、他の実験データが「0」でも除外してしまうので、チェックを入れない場合があります。

＊上記は弊社で解析を行う場合の設定例です。あくまで参考ですので、データに合わせて設定値の変更を行ってください。

【遺伝子セット登録について】
遺伝子セット登録のタブでは Search Option 等で検索したデータ群の Block、Row、Column の情報を保存し、遺伝子（or 抗体）セットとして登録させることができます。登録した検索条件を呼び出すこともでき、また登録してある情報に追加登録、あるいは削除登録することもできます。例えば、カテゴリー検索で「Cancer」を検索し、登録を行います。次に「Cell Cycle」を検索し、追加登録を選択すると、「Cancer」と「Cell Cycle」を合わせて遺伝子（or 抗体）セットの登録ができます。このように、お客様ご自身のオリジナルの遺伝子（or 抗体）セットを作成することができます。

新規登録　メイン画面で Select の項目に「○」がつけられた行の Block、Row、Column の情報を登録します。このとき遺伝子セットの色の設定を行うことができます。この色はスキャッタープロットを表示するときに反映されるので、他と識別しやすい色にしてください。また、スキャッタープロット図の各プロットをクリックすると赤に変色するようにプログラムされているので、なるべく赤色を登録することはお避けください。

追加登録　選択されている検索条件ファイルに、現在表示されている情報を追加登録します。
ただし、個々の遺伝子（or 抗体）情報は重複して登録されることはありません。

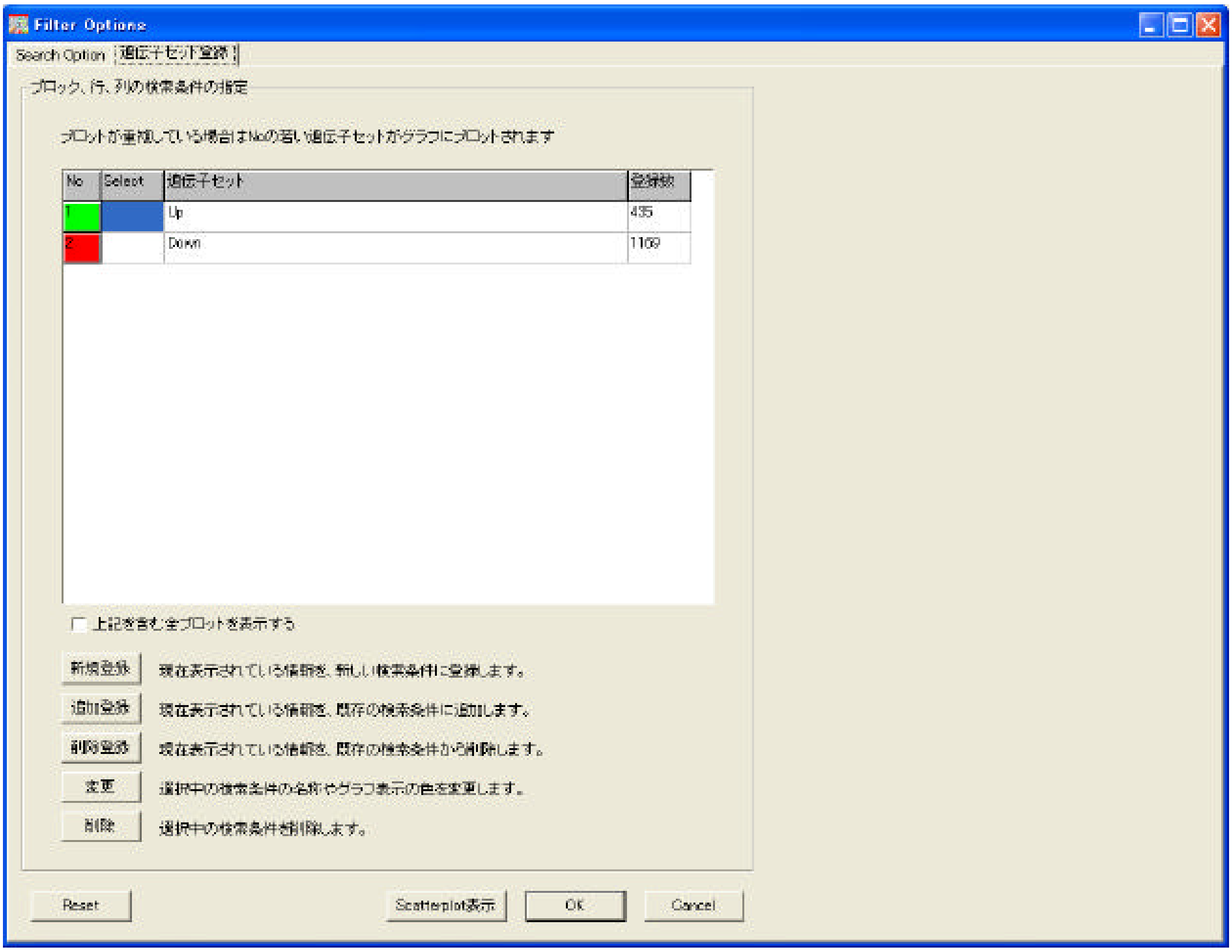
削除登録　選択されている検索条件ファイルから、現在表示されている情報を削除登録します。

変更　選択されている検索条件ファイルの名称やグラフ表示の色の変更を行います。

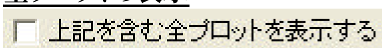
削除　選択されている検索条件ファイルを削除します。

遺伝子セットの並べ替え

:各遺伝子セットの行を右クリックすると、遺伝子セットの並べ替えを行うための選択肢が表示されます。
項目を選択することによって遺伝子セットの順番を変更することができます。複数の遺伝子セットを同時にScatterPlot表示するときに複数の遺伝子セットに重複して登録されている遺伝子がある場合、この番号が小さい方の遺伝子セットが優先されます。また、並べ替えを行うと、No.も変化します。



全プロットの表示



:このチェックボックスにチェックを入れると、全実験データのプロット上に、登録した遺伝子セットを色違いで同時に表示させることができます。ただし、[SearchOption]での検索条件の設定が残っていると、全遺伝子データのプロットが表示されません。全遺伝子データを表示させるには、一度検索条件のリセットを行い、メイン画面上に全データを表示させてください。この状態で前述の操作を行うと、全実験データと指定した遺伝子セットを同時にプロットさせることができます。

詳細はP.30【遺伝子セット登録画面からのScatterPlot】の項をご参照ください。

【検索対象ファイルの呼び出し】

Selectボックスより検索対象とする遺伝子(or抗体)セットを選択し、「OK」ボタンをクリックすると、選択した遺伝子(or抗体)セットに対応するデータを表示させることができます。遺伝子(or抗体)セットは複数選択できます。



OKボタン　：選択されている検索条件を実行します。



Cancelボタン　：選択されている検索条件を実行せず、Filter option画面を閉じます。



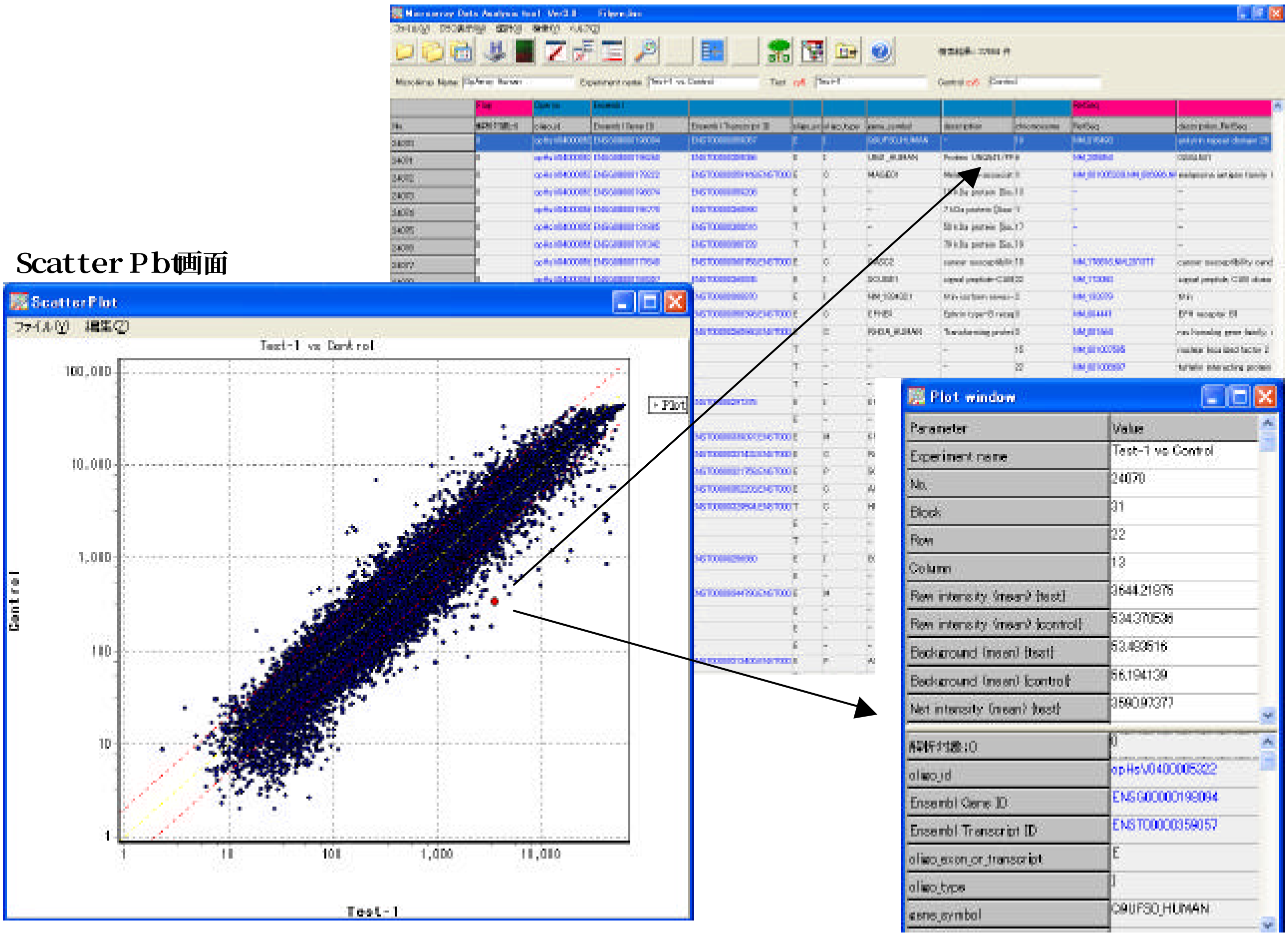
Scatterplot表示ボタン　：選択されている検索条件で遺伝子の絞込みを行ったデータに対するScatterPlotを表示させます。

5. ソフトウェアの操作　(Scatter Plot / Plot window)

Scatter Plot について

メイン画面の ScatterPlot 表示ボタンまたはメニューのグラフ表示で「ScatterPlot 表示」を選択すると下図のような「Scatter Plot」グラフが作成され、表示されます。ScatterPlot とは各スポットのテストサンプル側の Normalized intensity(または Per Array Normalized intensity)をX軸、コントロールサンプル側の Normalized intensity(または Per Array Normalized intensity)をY軸に分布した散布図です。本ソフトウェアでは対数表示でグラフを作成します。各スポットの intensity の値が近いほど、そのプロットは中心ライン（下図黄色ライン）に表示され、intensity の値に差があるほど中心ラインより離れた位置に表示されます。このプロットの分布により比較するサンプル間の遺伝子（or 抗体）発現状態の全体を直感的に判断することができます。なお、検索後のデータにおいても Scatter Plot グラフを表示させることができます。

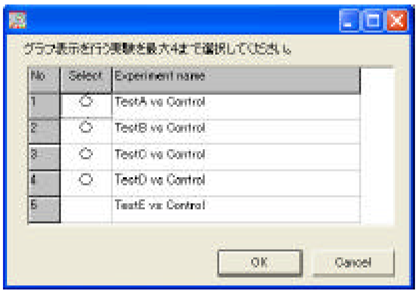
メイン画面



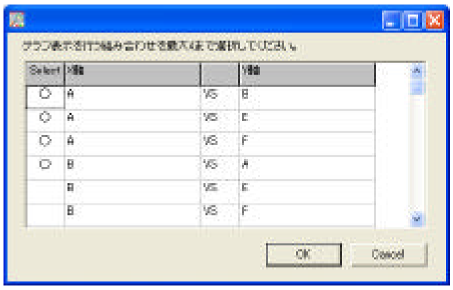
Scatter Plot のグラフ機能について

【表示させる組み合わせの選択】

複数比較、または Signal 比較データ解析の場合、表示させる実験データの組み合わせを選択する必要があります。例えば、テストサンプルとコントロールサンプルがペアになっている複数比較解析では、1 比較単位で指定を行いますが、この場合 X 軸とY 軸のサンプルは自動的に決定されています。しかしSignal 比較解析の場合は、比較を行う組み合わせが決められていないので、グラフ作成時に、組み合わせを決めなければいけません(次ページの図参照)。そのため Signal比較解析では、グラフ作成時に比較の組み合わせと、X 軸(テストサンプル)、Y 軸(コントロールサンプル)の指定を行います。なお、一度にグラフ表示できる組み合わせの数は 4 つまでです。



複数比較データ解析



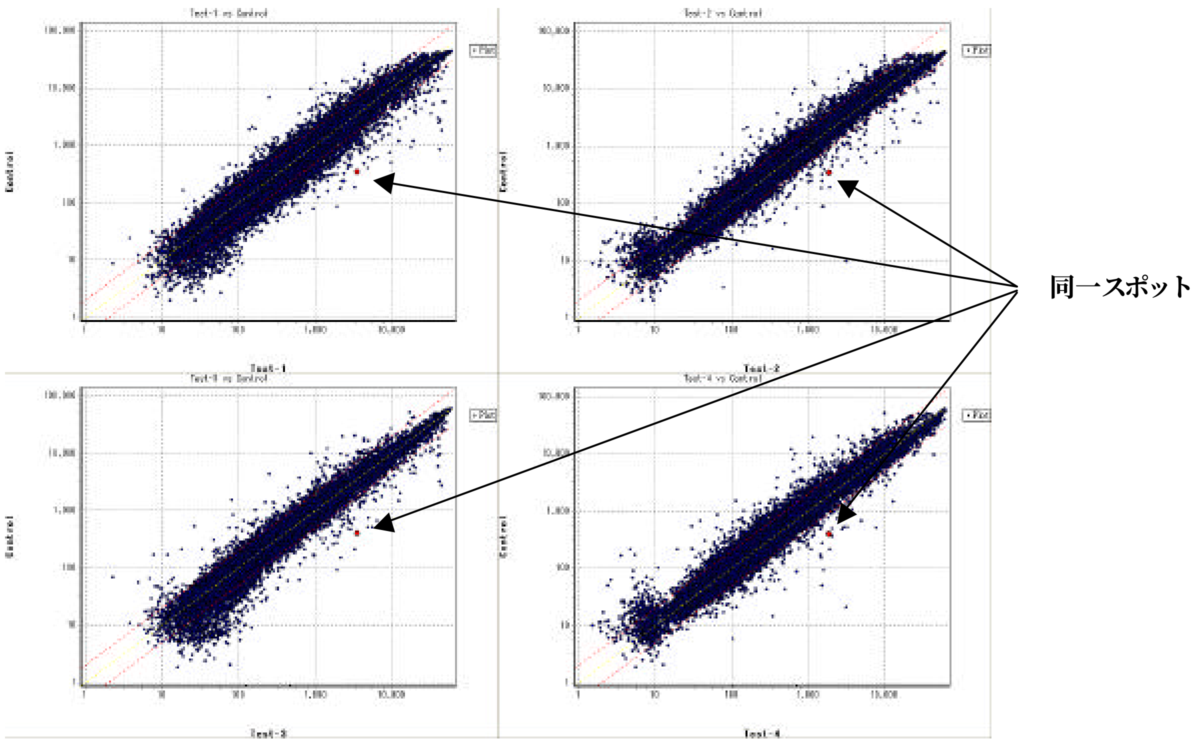
Signal 比較データ解析

【メイン画面とのリンク】

Scatter Plot の各プロットをクリックすると、選択されたプロットが赤いプロットに変ります。そして、メイン画面上の対応する遺伝子（行）が青色で選択され、行の先頭にスクロール表示されます。また、行が青色で選択されている状態で、マウスを右クリックあるいは、メイン画面のメニューの「選択」をクリックすると、選択された行の登録あるいは解除をすることができます。

複数比較、Signal 比較のデータを開いている場合は、最大 4 つまで同時に Scatter Plot を表示することができます。そして複数のグラフが表示されている場合、1つのグラフのプロットをクリックすると、他のグラフの対応するプロットも赤くなります。

また、Filter Options の遺伝子セットにチェックが入っており、複数の遺伝子セットが選択されている場合、Scatter Plot 表示については、それぞれの遺伝子セットが個別の色付きで表示されます。

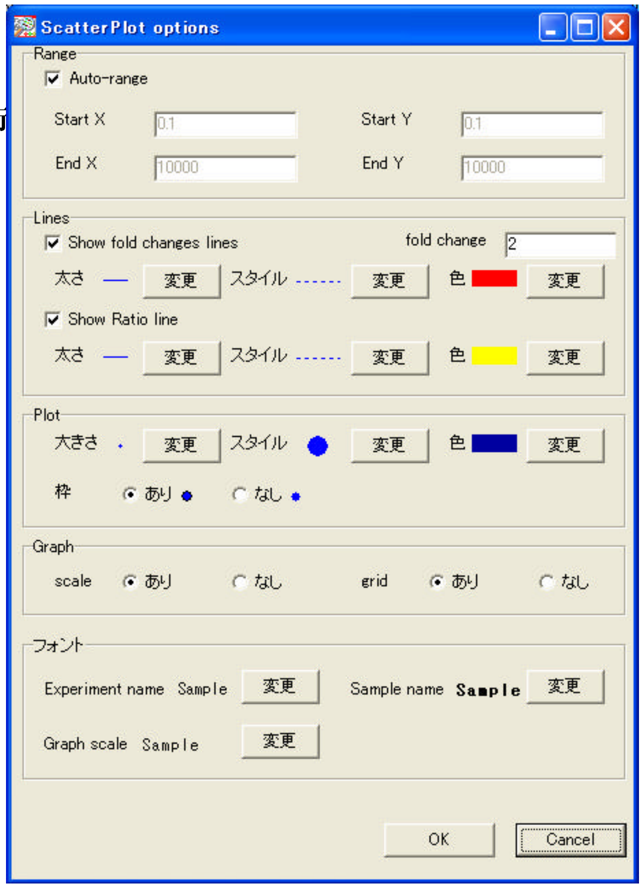


【Plot window 表示】

メイン画面の Plot window  アイコンをクリックすると、Scatter Plot 中の各プロット個別のデータを表示する Plot window 画面が現れます。別の Plot を選択すれば、表示されるデータも変更されます。また、表中の青文字で表示された遺伝子（or 抗体）アノテーションデータをダブルクリックすると、対応する公共データベースにハイパーリンクされます。但し、1 つのセルの中に複数のデータがある場合、あるいは「—」、「Blank」の場合にはハイパーリンク機能は対応していません。なお複数比較、Signal 比較のグラフの場合は、選択した実験名、あるいは比較名も表示されます。

【ScatterPlot options】

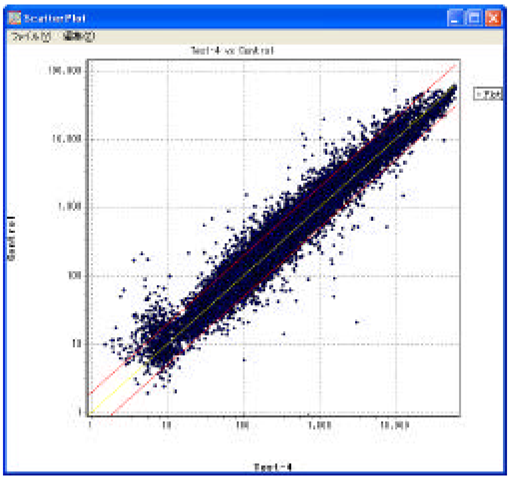
Scatter Plot 画面のメニューで編集を選択すると、Scatter Plot options 画面が表示されます。この画面では、グラフの表示や目盛りの各種設定を行います。



Range　Scatter Plot グラフのグラフレンジを設定します。「Auto-range」を選択すると、グラフのレンジ設定が自動となります。マニュアル設定の場合は、チェックボックスを解除し「Start X,End X」「Start Y,EndY」を入力し、設定を行います。設定を反映するには「OK」ボタンを押します。

Lines　ScatterPlot グラフのライン表示の詳細を設定します。

Show fold changes lines　fold change（発現 Ratio）ラインの有無とラインの太さ、スタイル、色、fold change の設定を行います。
fold change ラインとは発現 ratio を反映したラインであり、表示させたい発現 Ratio を入力し、ラインを表示させます。
下図は2倍の発現 Ratio のラインを引いた場合です（赤色）。



Show Ratio line　Ratio ラインの有無とラインの太さ、スタイル、色の設定を行います。 Ratio ラインとはグラフの中心ラインであり、発現 Ratio でいう「1倍」のラインになります。上図の黄色のラインが Ratio ラインになります。

Plot　Scatter Plot グラフの Plot 表示の詳細を設定します。Plot の大きさ、スタイル、色、枠の有無の設定を行います。

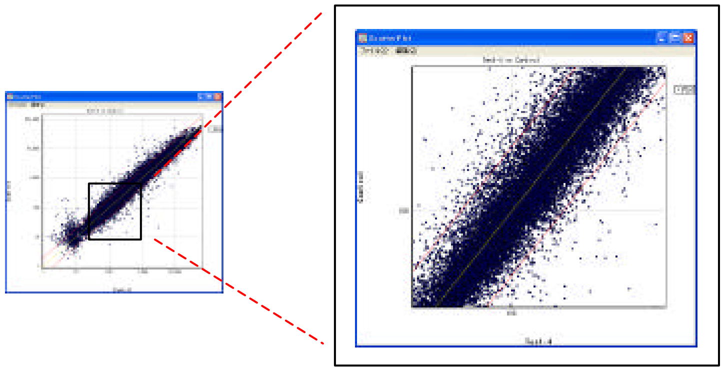
Graph　Scatter Plot グラフの目盛りの有無とグラフ背景のグリッド（格子）の有無の設定を行います。

フォント　Scatter Plot グラフに表示されるサンプル名と目盛りのフォント設定を行います。

OK　OKボタン　：変更した設定条件で、再度グラフが表示されます。

Cancel　Cancelボタン　：変更した設定条件を実行せず、Scatter Plot options 画面を閉じます。

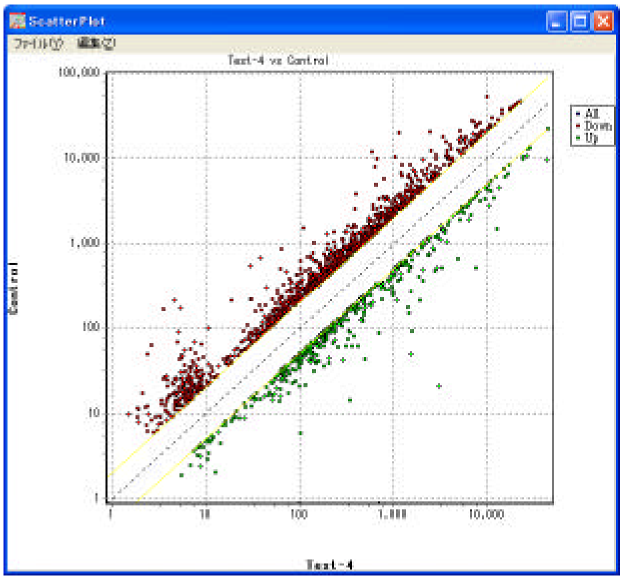
【Scatter Plot 拡大】
Scatter Plot グラフ上をマウス操作で右下方向に四角選択すると、選択された部分が拡大して表示されます。元に戻す場合は左上方向に四角を選択します。



【Scatter Plot 画像保存】
Scatter Plot 画面のメニューでファイルの「画像保存」を選択すると、表示されているScatter Plot グラフをBMP形式で保存できます。

【遺伝子セット登録画面からの ScatterPlot】
Filter Options の遺伝子セット登録画面で目的の遺伝子セットを選択します。
そこで画面下部の「Scatter Plot 表示」をクリックすると、選択した遺伝子セットの Scatter Plot が表示されます。

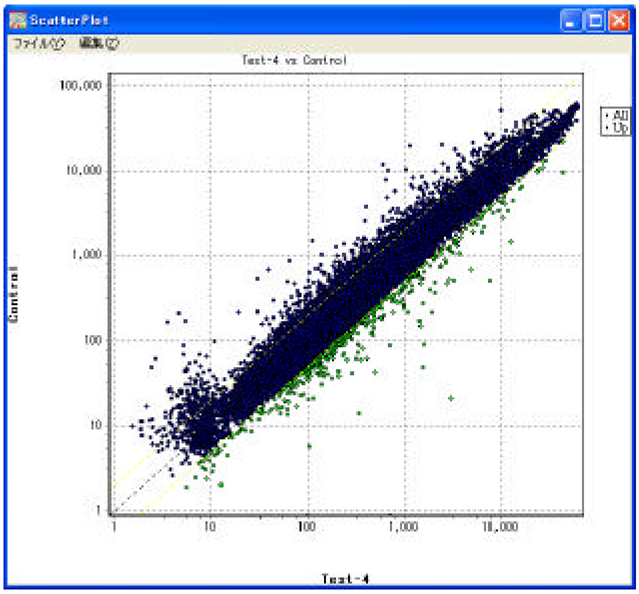
遺伝子セットを複数選択している場合、個々の遺伝子セットが色分けされてプロットされます。
プロットが複数の遺伝子セットに重複登録されている場合は、No.の小さい遺伝子セットの色で表示されます。



遺伝子セット登録画面における「上記を含む全プロットを表示する」のボックスにチェックを入れると、全データにおける任意の遺伝子セットの Scatter Plot の分布を見ることができます。
手順は、
① メイン画面上に全遺伝子データを表示させる。
② 「Filter Options」の遺伝子セット登録画面で、「上記を含む全プロットを表示する」のボックスにチェックを入れる。
③ 任意の遺伝子セットを選択する。
④ 「Scatter plot 表示」ボタンをクリックする。

このとき遺伝子セットで登録する色に注意してください。
全遺伝子データの色(この場合は青)と同じだと、プロットが識別しにくくなることがあります。

(ex 右図)
緑 :Test/Control の Ratio が 2 倍以上の遺伝子セット
青 :マイクロアレイ上の全遺伝子セット



6. ソフトウェアの操作 (統計検定)

統計検定について

複数または Signal 比較解析で 1 つ以上の実験データグループを設定している場合、グループ間またはグループ内の発現量に有意な差があるか統計検定を行うことができます。有意性の指標として P-value または FDR が全ての遺伝子ごとに出力されるので、これらの値を基準にして絞込みを行うことで、有意性をもつ遺伝子を容易に抽出することができます。

検定法の種類

本ソフトウェアで使用できる検定法には以下のものがあります。なお P-value の算出には、それぞれの実験について、複数比較解析では Normalized intensity (mean) {log2ratio}のデータを、Signal 比較解析では Per Gene Normalized intensity のデータを用いて計算を行っています。

①Student 's t-test

登録した実験データのグループ数が 2 つの場合に使用できます。それぞれのグループの発現量の平均値と標準偏差を比較し、帰無仮説「2 グループ間の発現量の平均値に差はない」に対する有意確率(P-value)を出力します。したがって、この P-value が小さい遺伝子ほど、グループ間の発現量に差があるといえます。なお P-value による絞込みでは、どちらのグループの方が発現量が大きいのかは分からないので、Ratio や Group Ratio で一方のグループで発現量が多い遺伝子のみに絞り込んでおき、その後 P-value を計算するといった手順が一般的に使われています。

②One-Way ANOVA

登録した実験データのグループ数が 2 つ以上の場合に使用できます。それぞれのグループ間の分散値とグループ内の分散値を比較し、帰無仮説「k(≧2)グループ間の発現量の平均値に差はない」に対する P-value を出力します。したがって、P-value が小さい遺伝子では、k グループいずれかのグループ間の発現量に差があるといえます。ただし 3 つ以上のグループがある場合に、この検定法ではどのグループ間に差があるのかまでは分からないので、絞込みの初期段階でグループ間発現量に有意差をもつ遺伝子を絞り込む目的で使用されます。

③One-Sample t-test

グループ間の発現量の違いを検定する他の 2 種類の手法と違い、1 つのグループ内の発現量が 0 からどの程度偏っているのかを検定します。つまりこの検定法は「平均値の差」ではなく、「差の平均値」＝変化量が同じかどうかを検定するので、この場合は P-value が小さいほどグループ内の発現量のばらつきが小さいといえます。グループ間発現量の違いを検定できないので、2 色法実験データが複数ある場合の、テストサンプルとコントロールサンプルの発現比に再現性のある遺伝子を絞込む目的で用いられています。またこの手法では、計算時に検定を行う実験グループを指定する必要があります。

オプション項目設定

検定処理時にオプション項目を設定することにより、外れ値の除去、多重性の補正などを行うことができます。これらの機能を有効にする場合は、各設定項目のチェックリストにチェックを入れてください。オプション項目には以下のものがあります。

①Ignored cells を除外する

P-value の計算時に、外れ値のあるデータ(Ignored cells が 1 になっているスポット)を除外して計算を行います。ただしグループ内に含まれている実験データ数が必要最低限の場合、この処理を行うことによって計算不可能な遺伝子がでてくる場合があります。計算不可能となった遺伝子は、P-value が 1 として出力されます。

②FDR (Benjamini and Hochberg)を計算する

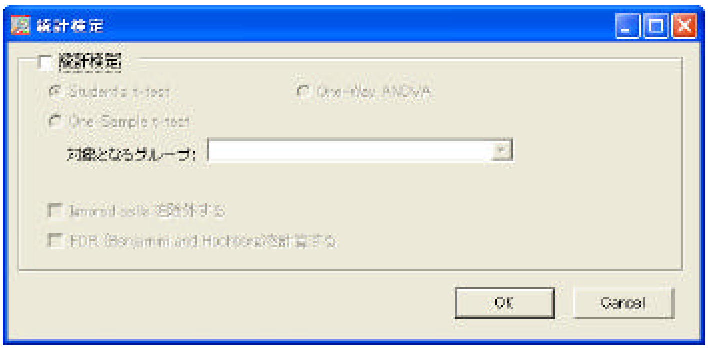
特定の現象が偶然起こり得る確率である P-value に対して、多重性を考慮し補正を行った値です。マイクロアレイ解析のように仮説が複数ある場合、つまり検定を行う遺伝子が複数ある場合は、帰無仮説を棄却するかどうかの有意水準は仮説が 1 個の場合よりも厳しくする必要があり、また仮説の数が増えるにつれ有意水準はより厳しくなっていきます。そのため検定を行う遺伝子の数(メイン画面上の遺伝子数)に応じて、P-value を補正しなければいけません。

FDR(False Discovery Rate)とは「有意と決定された仮説のうち本当は有意ではないもの」の比率を示しており、マイクロアレイ解析の多重性の補正で一般的に使用されています。データの性質は P-value と同じなので、この値が小さい遺伝子ほど有意であるといえます。なお本ソフトウェアでは、FDR の計算に Benjamini & Hochberg (1995)の手法を用いています。この手法では、検定時にメイン画面に表示されている遺伝子の件数を全仮説の数と見なして計算を行うので、計算を行う前に Filter Options での検索などで、メイン画面上の遺伝子数をできるだけ減らしておいた方が帰無仮説を棄却する有意水準は緩くなり、結果として FDR は小さくなります。

解析の手順

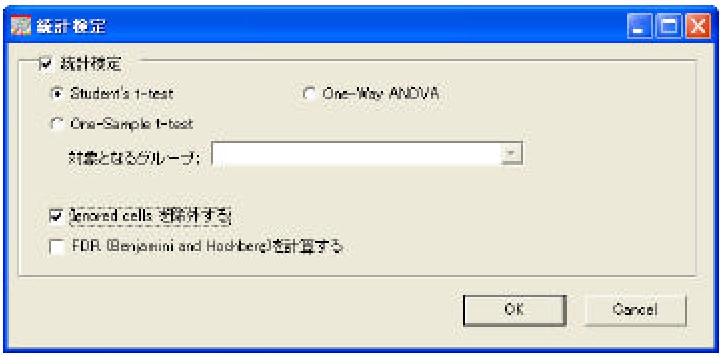
ソフトウェアを用いて実際に解析する場合の手順を示します。

①メイン画面でのアイコンをクリックすると、下のウインドウが開きます。ここで項目名「統計検定」の左側のチェックボックスにチェックを入れると、各種設定が行えるようになります。



*実験グループ設定を行っていない場合は、この機能は使用できません。

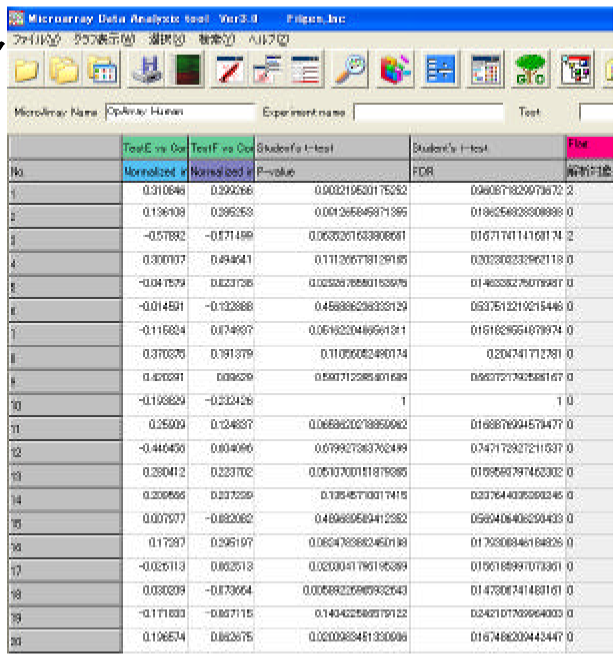
②設定したグループ数に合わせて使用する検定手法、およびオプション項目を選択してください。「One-Sample t-test」を使用する場合は、検定対象となるグループ名の選択が必要になります。



*FDRを計算する場合は、メイン画面上の遺伝子件数によって値が変わってくるので、検定を行う前にFilter Optionsなどで絞込みを行い、遺伝子件数を少なくしておいたほうが有意差が出やすくなります。(P.33 参照)

③「OK」ボタンをクリックすると、メイン画面の「P-value」の列に、「FDR (Benjamini and Hochberg)を計算する」にチェックを入れた場合は、「FDR」の列も同時に値が表示されます。また、それぞれの列の一番上の行に、使用した検定手法が表示されます。

ここで計算されたP-valueとFDRはFilter Options機能を使って検索を行うことができるので、有意差をもつ遺伝子のみを絞り込むことが可能です。
一般にP-value、FDRとも「0.05」以下になると有意性があるといわれています。



7. ソフトウェアの操作（Gene Ontology 解析）

Gene Ontology について

Gene Ontology（GO）とは生命現象に対する説明を階層的なツリー構造で表現したもので、ツリーの最上位階層で「molecular function」、「biological process」、「cellular component」の3つに大きく分類されています。この3つはそれぞれ分子機能、生命現象、細胞内局在に関するカテゴリーで、これらの下位クラスに「Apoptosis」、「cell division」などのより細かに分類されたカテゴリーが属しています。
そこで研究により機能が明らかになっている遺伝子には、その機能に関連したカテゴリー名(GO Term)が遺伝子ごと、生物種ごとにアノテーションとして付けられることになっています。また1つの遺伝子には複数の GO Term を付けることも許されており、関連付けられた GO Term を調べることでその遺伝子のもつ機能を一目で確認することができます。
＊この機能は Human、Mouse、Rat の生物種にのみ対応しています。

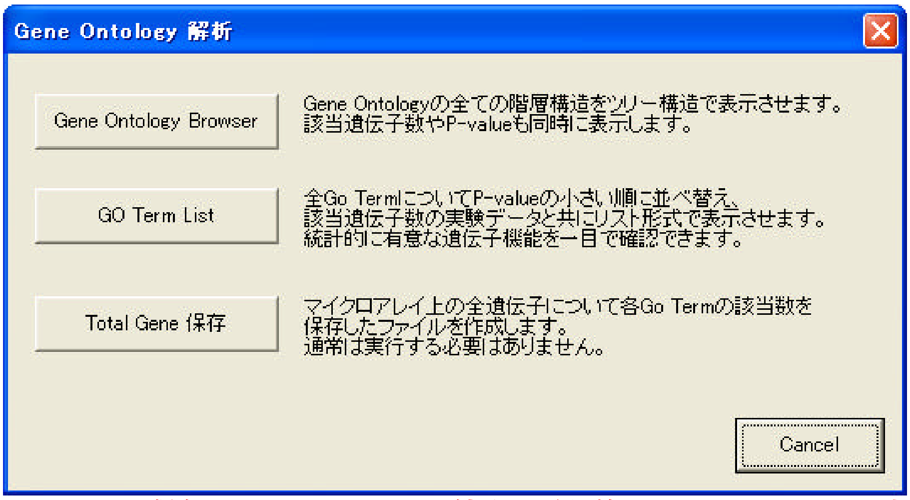
本ソフトウェアにおける GO 解析機能について

本ソフトウェアでは、通常の Gene Ontology 解析の他に、GO Term の中から重要なものだけを残した Gene Ontology の簡易版である GO Slim 解析もサポートしています。通常の GO 解析と違い、網羅性を犠牲にした分、GO Slim では計算時間が短く結果が簡潔で分かりやすいという特徴があります。
本ソフトウェアでは、実験データに対して自分で検索条件(例：ratio 2 倍以上、0.5 倍以下など)を指定し検索を行った後、検索した遺伝子群(メイン画面上の遺伝子)に対して、どのような GO Term をもつ遺伝子が多いのか統計的手法を用いて調べることができます。また、遺伝子の絞込みをせずに処理を行うことで、各 GO Term の全該当遺伝子データを見ることもできます。

【解析の手順】

① 実験データ（1比較、複数比較、Signal 比較のいずれでも可）を本ソフトウェアで読み込みます。
② 統計的な解析を行いたい場合は、Filter Options 機能などを使用し自分の興味のある遺伝子のみに絞り込みます。ただし該当遺伝子データのみを見たい場合は絞込みを行わずに、全データのままにします。
③ メイン画面の GO 解析アイコン をクリックし、データベース選択ウインドウを開きます。
ここで解析対象を「Gene Ontology」全部か、「GO Slim」か選択してください。
④ Browser 出力、または List 出力の選択を行います(下図)。また List 出力の場合は、List の形式も指定してください。(P.36、37 参照)

＊GO 検索結果画面を閉じてしまっても、Filter Options などの検索によってメイン画面の遺伝子数が変化しなければ、GO 計算結果は保持されます。つまり再度解析を実行するときは、計算時間は短く済みます。



＊Total Gene 保存ボタンをクリックすると、絞込みを行う前、すなわちマイクロアレイ上の全遺伝子についての各 GO Term に該当する遺伝子数を保存したファイルを作成します。このファイルは P-value を計算するのに必要ですが、納品の時点で添付されているので実行する必要はありません。
＊「Gene Ontology Browser」または「GO Term List」ボタンをクリックすると、計算時間に関する警告画面が表示され、ここで「OK」を選択すると計算を開始します。ここで表示される所要計算時間は、あくまで目安なのでお使いのコンピュータによっては計算時間が変化することがあります。

【解析結果データ】
本ソフトウェアでは解析結果として、各 GO Term につき以下のデータが出力されます。また、これ以降は全て「Gene Ontology」の解析を基準に説明しますが、「GO Slim」解析でも共通の内容となります。

(1)Changed Genes
Filter Options などを用いて絞り込んだ遺伝子、すなわち GO 解析アイコンをクリックした時点で、メイン画面上に表示されている全遺伝子の中で、各 GO Term に関連付けられている遺伝子数を表しています。

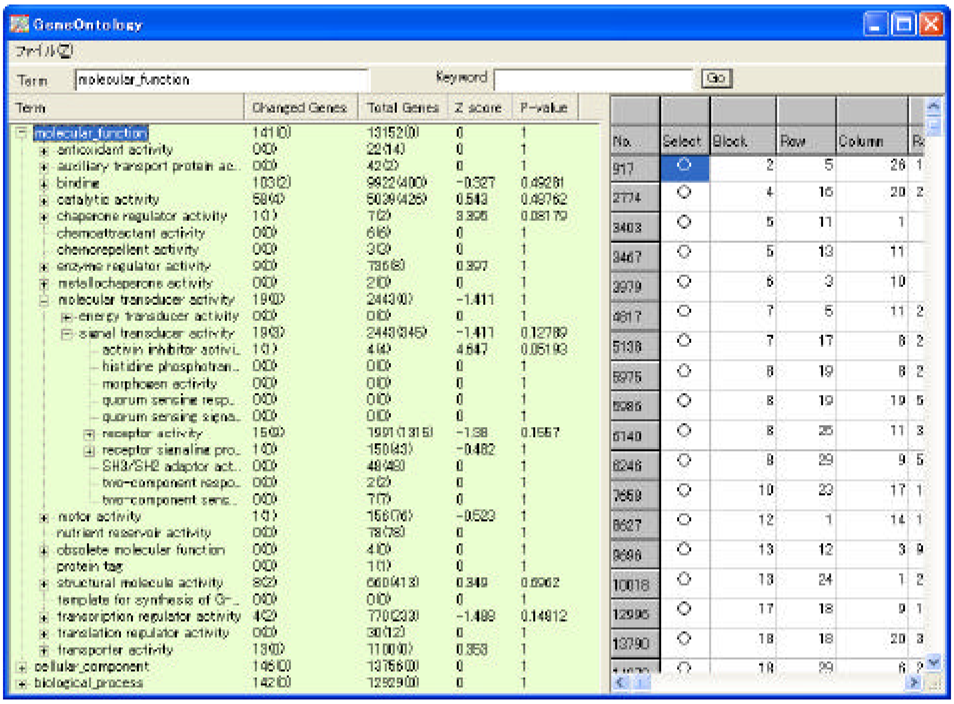
(2)Total Genes
絞込みを行う前の遺伝子、すなわち各種マイクロアレイに搭載されている全遺伝子の中で、各 GO Term に関連付けられている遺伝子数を表しています。

(3)Z-score
Changed Genes と Total Genes の比率から、Changed Genes の数が平均値からどれだけ外れているかを示した値です。この値がプラスのときは平均値より大きい、マイナスのときは平均値より小さいことを表し、絶対値が大きくなるにつれ平均からのずれが大きいことになります。基本的にこの値が大きい GO Term ほど、有意な機能であるといえます。

(4)P-value
Changed Genes と Total Genes の比率から、帰無仮説「Changed Genes と Total Genes の間で、特定の GO Term をもつ遺伝子ともたない遺伝子の比率に差はない」に対する有意確率のことです。したがって、この値が小さいほど Changed Genes と Total Genes の間でその GO Term をもつ遺伝子の比率に違いがあるということになり、有意な機能であるといえます。ただし、Z-score と違って比率のずれの性質(平均値より多いか少ないか)を識別することはできないので、有意な GO Term の抽出には Z-score と組み合わせて検索を行います。

【出力形式】
本ソフトウェアでは検索した Gene Ontology について 2種類の表示方法があります。

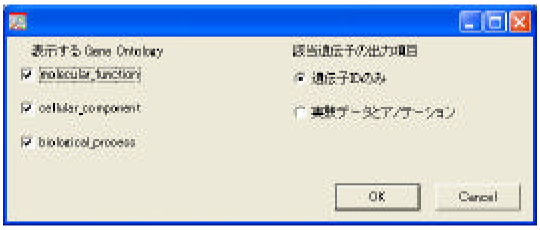
①Gene Ontology Browser(GO Slim Browser)について
全ての GO Term についてツリー構造で表示させます。画面は大きく分けて 2つに分かれており、左側に Gene Ontology のツリー構造と各 Term の解析結果データ、右側に該当する GO Term をもつ遺伝子の実験データが表示されます。最初の状態では GO Term は最上位の 3つしか表示されていませんが、各 GO Term の先頭の[+]アイコンをクリックしていくごとに、下位階層が表示されていきます。そして各 GO Term をダブルクリックすると右側の領域に対応する遺伝子の実験データが表示されます。また、ウインドウ上部にある「Keyword」入力欄に任意の単語を入力し、「Go」ボタンをクリックするとその単語を含む GO Term の位置にジャンプします。Term が複数ある場合は、「Go」ボタンをクリックするたびに次の Term の位置に飛びます。



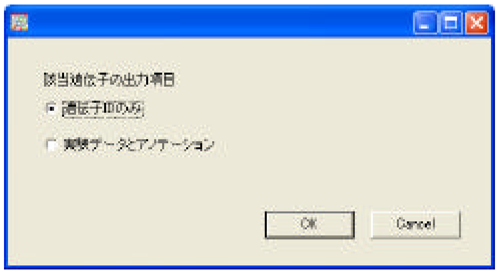
②GO Term List について

GO Term をP-valueの小さい順に並び替え、リスト形式で出力します。P-valueは統計的な有意確率を表しているので、この値が小さい順に並び替えることによって、統計的に有意なGO Term を一目で確認できます。このリストは出力時に様々な設定を行うことで、カテゴリーの絞込みや、該当遺伝子の実験データをつけることができます。

GO Term List 設定画面



GO Slim List 設定画面

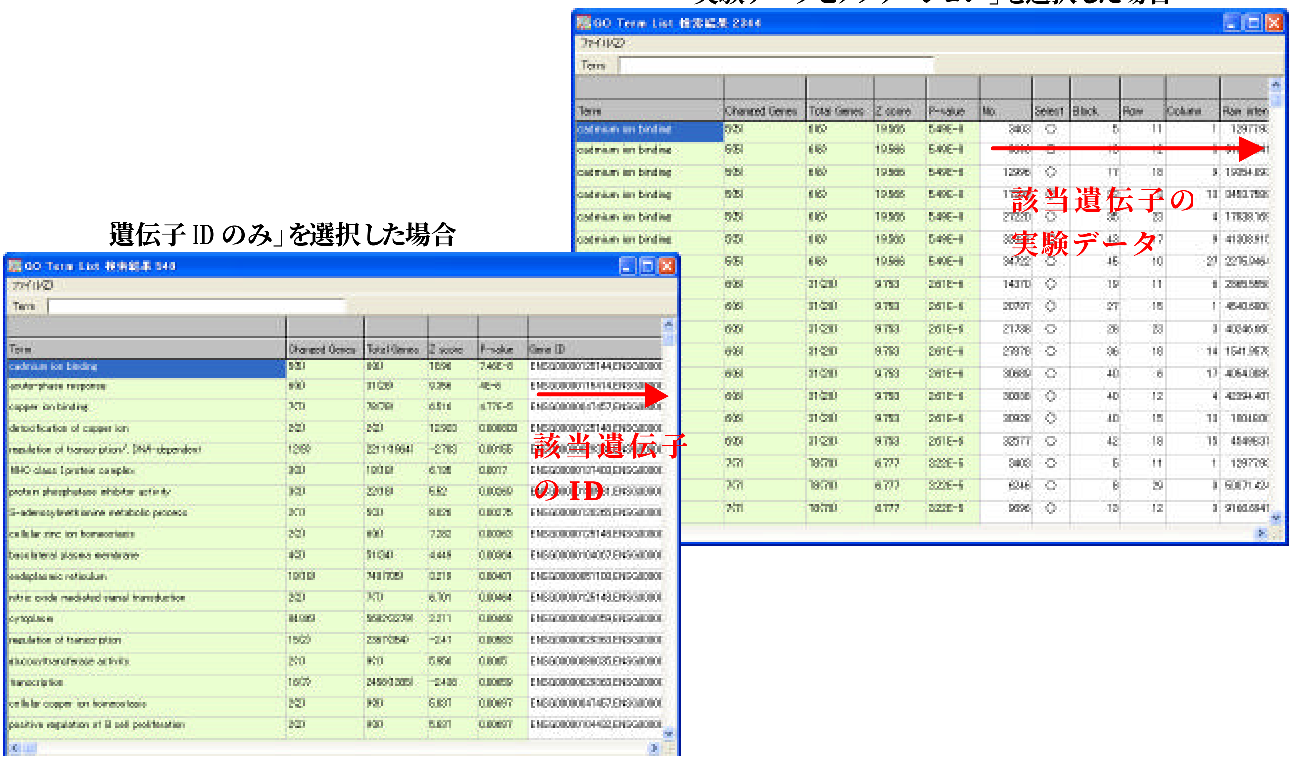


【表示するGene Ontology について】

上図のように、Gene Ontology 解析とGO Slim 解析では、設定を変更できる項目が違います。Gene Ontology 解析で出力するGO Term Listでは、リスト表示するTerm に対して「molecular_function」、「cellular_component」、「biological_process」のカテゴリー別に絞り込むことができます。その一方、GO Slim 解析によるリストではカテゴリーの絞込みは行えません。

「実験データとアノテーション」を選択した場合

「遺伝子 ID のみ」を選択した場合



【該当遺伝子の出力項目について】

各 GO Term に対応する該当遺伝子データは、その表示項目および表示形式を選択することができます。例えば、リスト内の各GO Term には、それぞれ対応する遺伝子のマイクロアレイ実験データが１つ以上ありますが、このとき該当遺伝子の遺伝子IDのみを表示させるか、または intensity や ratio などの実験データを全て表示させるか選択することができます。「遺伝子IDのみ」を選択した場合、各 GO Term の該当遺伝子の ID (Ensembl ID、RefSeq ID など) がカンマ区切りで１行にまとめて表示されます。対して「実験データとアノテーション」を選択した場合は、各 GO Term の該当遺伝子の実験データが全て表示されます。この場合、１遺伝子の実験データだけで１行使用するので、１つの GO Term に対して該当遺伝子のデータが複数行にわたって表示されます。なお、これらのリストはテキスト形式で保存することができます。

[計算アルゴリズム]

各 GO Term における遺伝子数の数え方

遺伝子と GO Term は各遺伝子の ID (Ensembl ID、RefSeq ID など) を用いて対応付けしております。つまり違う遺伝子であっても ID が同一であったり、また同じ遺伝子であっても ID が違う場合などで、GO Term の該当遺伝子数が変化します。このため、表中の該当遺伝子数と、該当実験データの数が合わなくなることがあります。

階層構造の数え方

本ソフトウェアでは全 GO Term について、リスト中の該当遺伝子数を計算します。ただし、Gene Ontology は階層構造をもっているため、各 GO Term の上位 -下位関係を考慮しないといけません。このため各 GO Term の該当遺伝子数(Changed Genes、Total Genes 両方とも)は、階層構造を考慮にいれて数えた数値と考慮していない数値の 2 種類があり、それぞれ解析結果画面上でカッコなし、カッコありの数値として表示されています。

例えばある GO Term における Changed Genes の項目が「15(4)」と表示されている場合、これはこの GO Term に 4 つの遺伝子が対応付けられており、その下の階層の GO Term に対応付いた遺伝子を合わせると 15 個の遺伝子が対応付けられていることを表しています。

Z-score について

各 GO Term の「Changed Genes」の値が、マイクロアレイ上の全遺伝子数に対するメイン画面上の遺伝子数の比率から計算した各 GO Term の該当遺伝子数の期待値から、どれだけ外れているかを示した値です。各 GO Term の「Changed Genes」と「Total Genes」、さらにその GO Term の最上位 GO Term における「Changed Genes」と「Total Genes」の値、これら4 つの数値を用いて計算を行います。また計算は全て階層構造を考慮した方の数値(カッコなしの数値)で行っています。

この値がプラスのときは期待値より大きい、マイナスのときは期待値より小さいことを表し、絶対値が大きくなるにつれ期待値からのずれが大きいことになります。したがって Z-score が大きい GO Term では、絞込み後の遺伝子リストに該当遺伝子が通常より多く含まれていることになり、有意な機能であるといえます。

P-value の求め方

本ソフトウェアでは各 GO Term における P-value(有意確率)の計算には Two tailed Fisher 's exact test を採用しています。この検定法では各 GO Term の「Changed Genes」と「Total Genes」、さらにそれぞれをその GO Term の最上位 GO Term における「Changed Genes」と「Total Genes」から引いた値、これら4 つの数値を用いて計算を行います。また計算は全て階層構造を考慮した方の数値(カッコなしの数値)で行っています。

Two tailed Fisher 's exact test は 2 群の比率の差の検定を行うための検定法です。本ソフトウェアでは遺伝子の絞込みを行う前と後で、各 GO Term をもっている遺伝子ともたない遺伝子の比率の差を検定し、帰無仮説「絞込みを行う前と後で、特定の GO Term をもつ遺伝子ともたない遺伝子の比率に差はない」に対する有意確率を求めることができます。つまり有意確率である P-value が小さいほど(0.05 以下)、絞込みの前後でその GO Term をもつ遺伝子の比率に違いがあり、その結果絞込み後の遺伝子で大きく変化している生物学的機能であるといえます。

＊Fisher 's exact test は 2 群の比率の差のみに注目しているので、データの絞込み後に有意に多い GO Term だけではなく有意に少ない GO Term も同時に検出してしまいます。このため有意な GO Term を抽出するには、P-value だけではなく Z-score がプラスの値になっているかどうかも確認してください。

Gene Ontology 情報のアップデートについて

納品の時点では、Gene Ontology 情報は最新のものを使用しています。その後のアップデートが必要な場合は、別途お問い合わせください。

8. ソフトウェアの操作 （Pathway解析）

Pathway について

Pathway(パスウェイ)とは、タンパク質や低分子化合物などによる細胞内相互作用情報を、機能や経路ごとに分類し、図の形式で表したものです。本ソフトウェアでは、パスウェイデータとして「GenMAPP (http://www.genmapp.org/)」を使用し、マイクロアレイに搭載されている各遺伝子と、GenMAPP で分類されている各パスウェイデータとの対応を調べることができます。この機能を用いることによって、解析遺伝子群の相互作用情報を直感的に捉えることができます。
＊この機能は Human、Mouse、Rat の生物種にのみ対応しています。

本ソフトウェアにおける Pathway 解析機能について

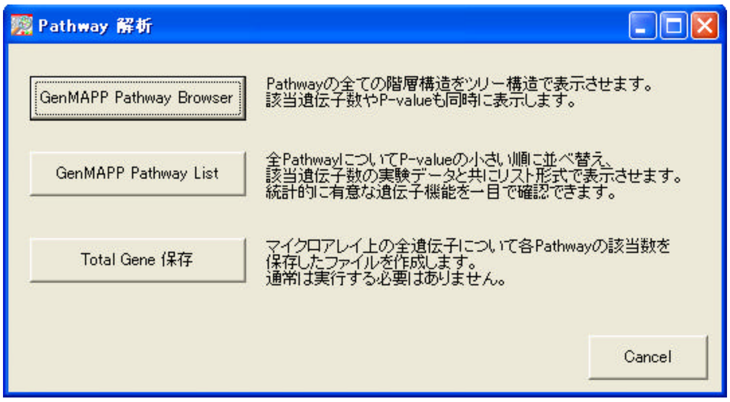
GenMAPP のパスウェイデータは大まかに 4 つに分かれており、そこからさらに具体的なパスウェイに分類されています。ただし、生物種によってこれらの分類が異なることがあります。

1. Contributed ：既知の分子間相互作用情報をパスウェイごとに分類し、図で表したものです。GenMAPP のパスウェイデータというと、この「Contributed」を指すことがほとんどです。
2. GO samples ：Gene Ontology での分類をより簡潔に表したものです。分類の数は多いのですが、この図には「Contributed」のような相互作用情報までは記されていません。
3. KEGG converted ：KEGG (Kyoto Encyclopedia of Genes and Genomes) によるパスウェイデータを独自に編集したものです。「Contributed」と同様、分子間相互作用を表したパスウェイ図があります。
4. Tissue-specific ：特定の臓器で発現している遺伝子ごとに分類したものです。パスウェイ図は「GO samples」と同様、分子間の相互作用までは記されていません。

Pathway 解析では Gene Ontology 解析と同様に、実験データに対して自分で検索条件(例：ratio 2 倍以上、0.5 倍以下など)を指定し検索を行った後、検索した遺伝子群(メイン画面上の遺伝子)に対して、どのような Pathway をもつ遺伝子が多いのか統計的手法を用いて調べることができます。また、遺伝子の絞込みをせずに処理を行うことで、各 Pathway の全該当遺伝子データを見ることもできます。

【解析の手順】

① 実験データ(1比較、複数比較、Signal 比較のいずれでも可)を本ソフトウェアで読み込みます。
② 統計的な解析を行いたい場合は、Filter Options 機能などを使用し自分の興味のある遺伝子のみに絞り込みます。ただし該当遺伝子データのみを見たい場合は絞込みを行わずに、全データのままにします。
③ メイン画面の Pathway 解析アイコン をクリックし、出力形式の選択画面を開きます。。
④ Browser 出力、または List 出力の選択を行います。また List 出力の場合は、List の形式も指定してください。



＊Total Gene 保存ボタンをクリックすると、絞込みを行う前、すなわちマイクロアレイ上の全遺伝子についての各 pathway に該当する遺伝子数を保存したファイルを作成します。通常は実行する必要はありません。

【解析結果データ】

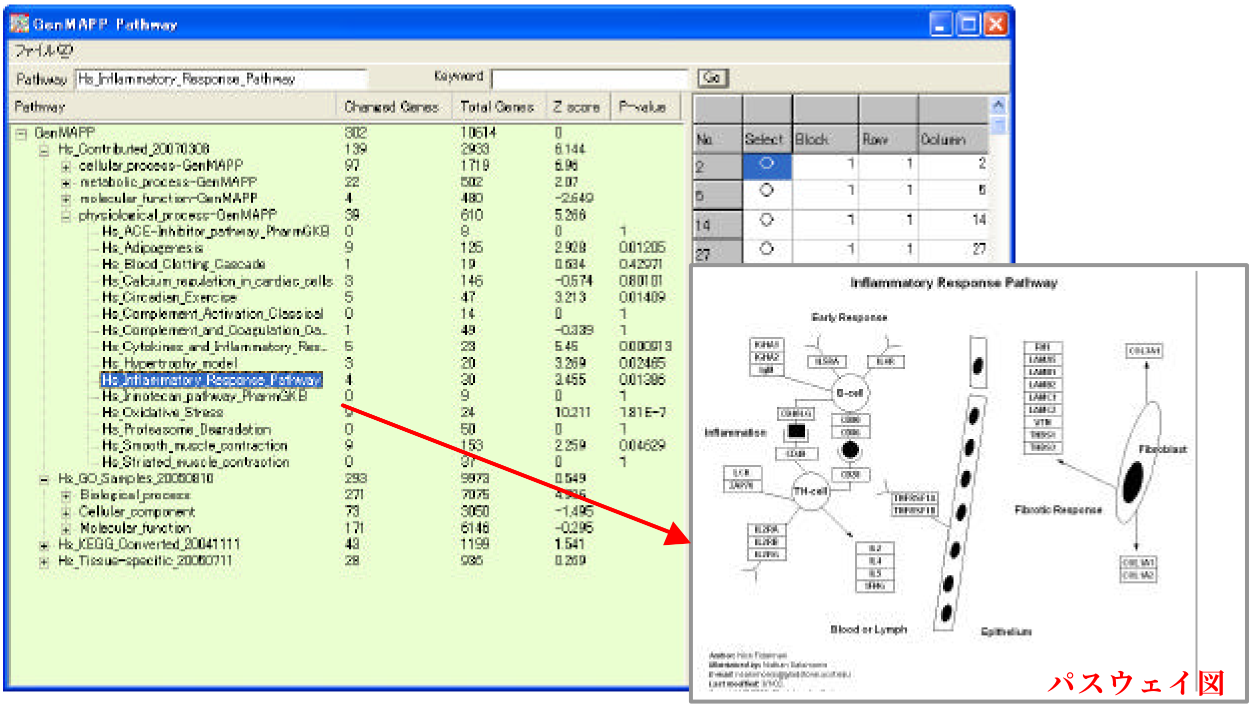
Pathway 解析の解析結果の見方は、基本的に Gene Ontology 解析と同じです。「Changed Genes」、「P-value」などについては Gene Ontology 解析のページ (P.36)をご覧下さい。

【出力形式】

Gene Ontology 解析と同様、検索した Pathway 情報について 2種類の表示方法があります。

①GenMAPP Pathway Browser について

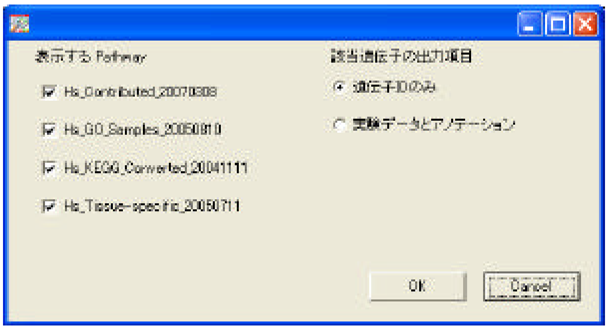
全てのパスウェイをツリー構造で表示させます。画面は大きく2つに分かれており、左側に GenMAPP Pathway のツリー構造と各パスウェイの解析結果データ、右側に該当するパスウェイをもつ遺伝子の実験データが表示されます。最初の状態ではパスウェイは最上位の「GenMAPP」しか表示されていませんが、各パスウェイ名の先頭の[+]アイコンをクリックしていくごとに、下位階層が表示されていきます。そして各パスウェイ名をダブルクリックすると右側の領域に対応する遺伝子の実験データが表示されます。また、ウインドウ上部にある「Keyword」入力欄に任意の単語を入力し、「Go」ボタンをクリックするとその単語を含むパスウェイの位置にジャンプします。パスウェイが複数ある場合は、「Go」ボタンをクリックするたびに次のパスウェイの位置に飛びます。



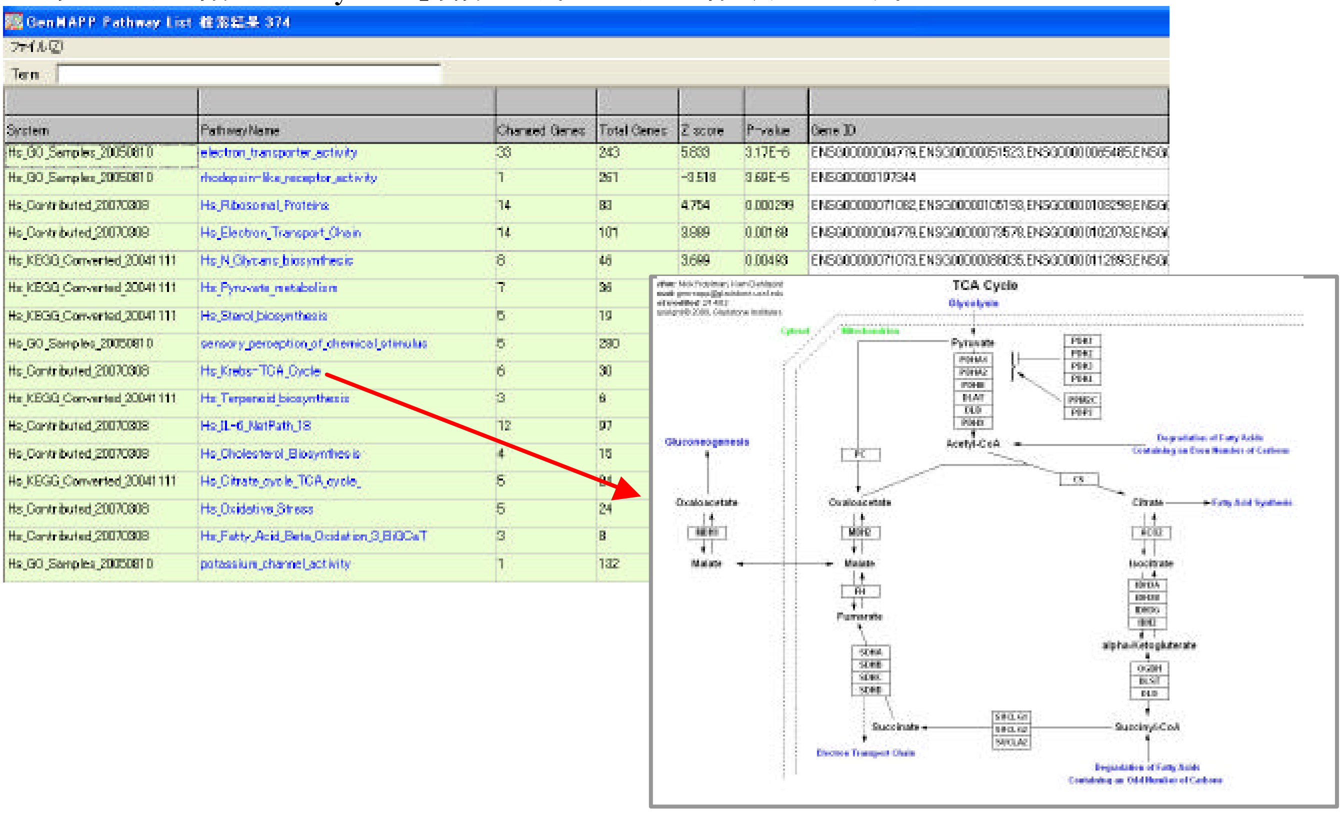
Gene Ontology Browser との違いは、GenMAPP Pathway の上位階層の各カテゴリーが、パスウェイの名前ではなく、単なる分類のためのクラス名だということです。そのため Pathway 解析では、「Contributed」や「Biological process」などの上位カテゴリーでは P-value の計算を行っていません。最下位層の正式なパスウェイカテゴリーでのみ、P-value を計算しています。
また各パスウェイ名を右クリックし、「図表示(Z)」というメニュー項目をクリックすると、そのパスウェイの図を表示することができます。

②GenMAPP Pathway List について

各パスウェイを P-value の小さい順に並び替え、リスト形式で出力します。P-value は統計的な有意確率を表しているので、この値が小さい順に並び替えることによって、統計的に有意なパスウェイを一目で確認できます。GO Term List と同様、出力時に様々な設定を行うことで、カテゴリーの絞込みや、該当遺伝子実験データの表示方法を選択することができます(下図)。該当遺伝子の出力項目については、Gene Ontology 解析のページ (P.37)をご覧下さい。



GenMAPP Pathway List では、GenMAPP Pathway Browser と同様、各パスウェイ図へのリンクが貼ってあります。「Pathway Name」列の青字で表示されている各パスウェイ名をダブルクリックしてください。なお、リスト中の「System」項目は各パスウェイの上位カテゴリー名、「Pathway Name」項目が正式なパスウェイ名を表しています。



[計算アルゴリズム]

各パスウェイにおける遺伝子数の数え方

Gene Ontology 解析と同様、遺伝子とパスウェイ(Pathway Name)は各遺伝子の ID (Ensembl ID、RefSeq ID など) を用いて対応付けしております。つまり違う遺伝子であっても ID が同一であったり、また同じ遺伝子であっても ID が違う場合などで、パスウェイの該当遺伝子数が変化します。このため、表中の該当遺伝子数と、該当実験データの数が合わなくなることがあります。また、Pathway Name より上位のカテゴリー(「Hs_Contributed_20070308」、「cellular_process-GenMAPP」など)の該当遺伝子数は、下位のパスウェイに含まれる全ての遺伝子数を、重複を除いて足した値になっています。

Z-score について

Gene Ontology 解析の場合と同様です。詳細は Gene Ontology 解析のページ (P.38)をご覧下さい。

P-value の求め方

Gene Ontology 解析の場合と同様です。詳細は Gene Ontology 解析のページ (P.38)をご覧下さい。

GenMAPP Pathway 情報のアップデートについて

納品の時点では、GenMAPP Pathway 情報は最新のものを使用しています。その後のアップデートが必要な場合は、別途お問い合わせください。

9. ソフトウェアの操作 （External Filter 機能）

External Filter について

自分で作成した遺伝子リストのような外部データと、本ソフトウェア上のマイクロアレイデータの統合を行うための機能です。2 つのデータの対応付けには、それぞれ対応付けのためのアノテーション項目列を指定し、その 2 列についてキーワード検索を行うことによって対応させます。

外部データ=Oligo_ID　メインデータ=oligo_id　検索結果 12件

ファイル(Z)

Oligo_ID	Ensembl Gene ID	Gene Symbol	No.	Select	Block	Row	Column	Raw intensity	Raw intensity	Background (	Background (
H300012007	ENSG00000060237	PRKWNK1	1899	○	3	13	9	559.933036	1232.026786	46.076923	53.556777
H200013401	ENSG00000060749	NP_079050	7970	○	11	6	5	823.921875	1598.386161	53.897436	60.912088
H300018481	ENSG00000061273	HDAC7A	21544	○	28	15	25	234.765625	217.350446	47.047619	52.578755
H300016373	ENSG00000061938	ACK1_HUMAN	10558	○	14	15	1	316.167411	647.247768	50.67033	60.010989
H300011438	ENSG00000062194	NP_075064	32435	○	42	13	8	446.553571	604.506696	56.014652	58.893773
H200007926	ENSG00000067369	TP53BP1	35328	○	46	4	12	1210.779018	2256.772321	56.556777	60.212454
H200006046	ENSG00000067829	IDH3G	34524	○	45	3	18	438.323661	1129.276786	46.791209	53.362637
H300017214	ENSG00000197124	NP_149973	16832	○	22	15	11	90.803571	144.029018	50.443223	63.589744
H200009684	ENSG00000068120	COASY	7160	○	10	5	5	325.194196	673.979911	50.058608	59.52381
H200014246	ENSG00000068308	NP_060072	26771	○	35	6	14	409.671875	427.859375	56.761905	56.81685
H300011281	ENSG00000068724	TTC7A	15202	○	20	13	1	356.127232	476.328125	51.564103	57.186813
H300022604	ENSG00000068781	T2AY_HUMAN	34133	○	44	18	5	1548.421875	917.834821	52.593407	47.465201

←外部データ　　　マイクロアレイデータ→

外部データのフォーマット

対応させる外部データは以下のフォーマットに従うように作成してください。

① タブ区切りのテキストファイルとする。
② 1 行目にはそれぞれの列名を記入し、2 行目以降を実際のデータとする。
③ 1 列目には必ず行番号を記入し、この列は検索に使用しない。

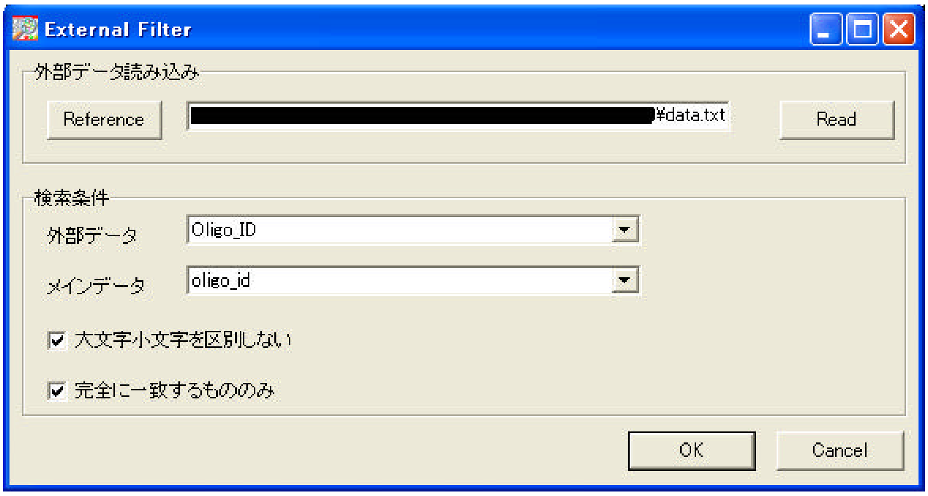
外部データ作成例

	A	B	C	D	E
1	検索データ	検索oligo_id	検索oligo_id	検索EnsemblGeneID	検索transcript_id
2	1	H200000226	H200000226	ENSG00000004478	ENST00000001008
3	2	H300015533	H300015533	ENSG00000000457	ENST00000158166
4	3	H300016407	H300016407		
5	4	H300013738	H300013738		
6	5	H200016156	h200016156	ENSG00000005073	ENST00000006015
7	6	H300010851	h300010851	ENSG00000005102	ENST00000318579
8	7	H200010569	h200010569	ENSG00000196712	ENST00000356175
9	8	H300018387	h300018387	ENSG00000005379	ENST00000343736
10	9	H200017528	h200017528	ENSG00000010017	ENST00000011619
11	10	H200002827	h200002827	ENSG00000010165	ENST00000362019
12	11	H200002595	h200002595		
13	12	H200019995	H200019995		

External Filter 解析の手順

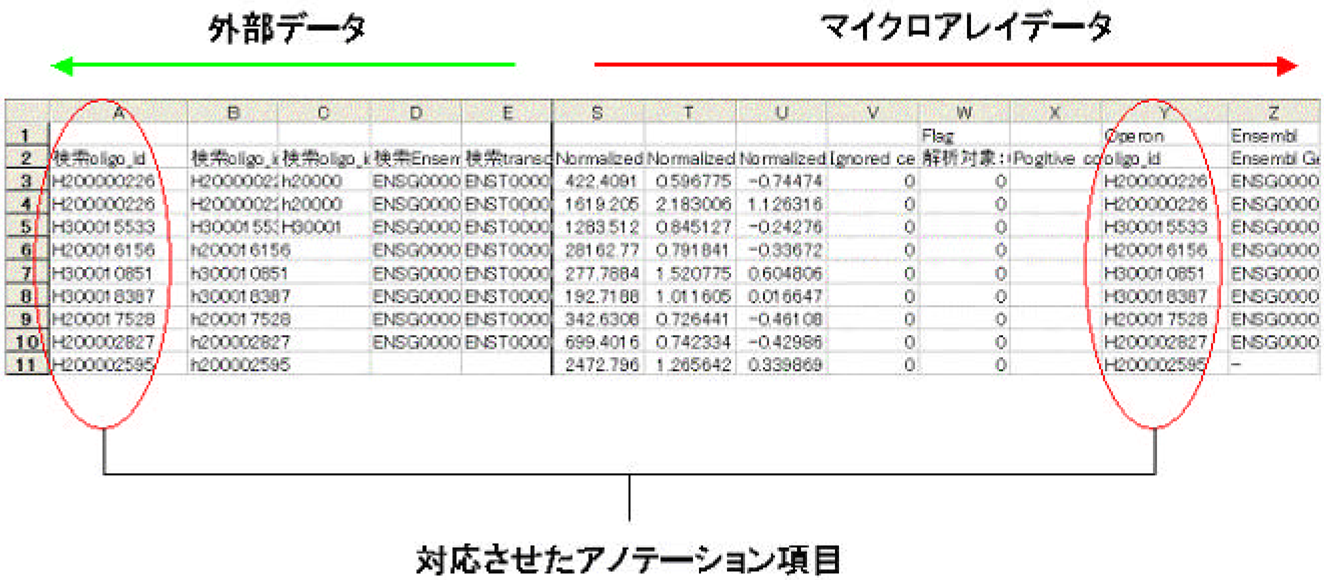
① 対応させたい外部データを準備する。

② メイン画面の External Filter アイコン をクリックし、外部データ読み込みウインドウを表示させる。
③ 外部データ読み込みウインドウの「Reference」ボタンをクリックし、外部データの指定を行う。
④ 「Read」ボタンをクリックし、外部データ、マイクロアレイデータそれぞれの列情報を読み込む。
⑤ 「検索条件」の項目で、外部データとマイクロアレイデータを対応させるキーワードを含む列をそれぞれについて指定し、状況によって大文字小文字の区別、キーワードの完全一致、部分一致の指定を行う。



外部データ読み込みウインドウ設定例

【外部データ？ マイクロアレイデータ統合例】



10. ソフトウェアの操作　（クラスタリング用データ出力）

クラスタリング用データ出力について

本ソフトウェアではStanford大学Eisen,Labからダウンロードできる統計解析フリーソフトウェア「Cluster and TreeView」のファイルフォーマット用にデータを出力させることができます。

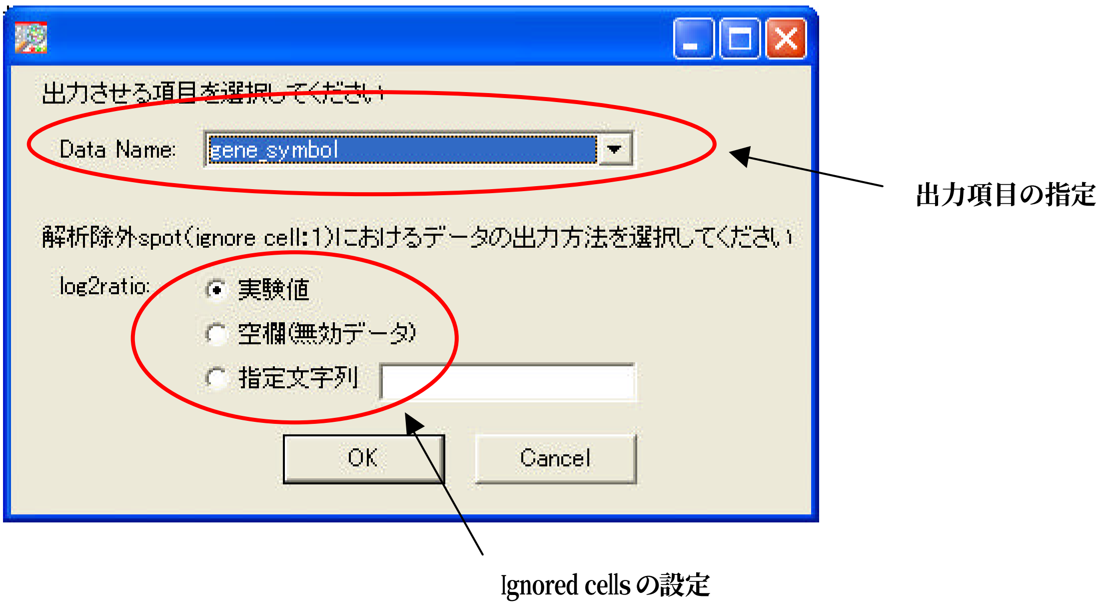
【データ出力の方法】

①クラスター解析を行いたいデータをメイン画面に表示させ、メイン画面のクラスタリング用データ出力アイコンをクリックします。
②プルダウンメニューからデータと共に出力させたい項目、さらにIgnored cellsのデータの処理を選択します。項目は1つしか選択できません。選択後、「OK」ボタンを押し、ファイルを保存します。

【Ignored cellsの設定】

Ignored cellsの設定がされているスポットのデータはノイズが混ざっているので、クラスタリングのときにデータの削除、または変換をする必要があります。本ソフトウェアでは、データ出力の際、Ignored cellsスポットのデータをどう扱うかを指定できます。

①実験値　　:変換をせず、そのままのデータを出力します。
②空欄　　　:データなしとして出力します。この場合、Ignored cellsスポットのデータは計算に使用されません。
③指定文字列 :特定の文字列に変換します。このとき変換後のデータが数値の場合は計算に使用され、文字列の場合は計算されません。



11. Cluster and TreeView の使い方について

Stanford大学 Eisen,Labからダウンロードできる統計解析フリーソフトウェア「Cluster and TreeView」を用いることで各種クラスタリング解析を行うことができます。ここでは、「Microarray Data Analysis Tool」で出力されたデータファイル(.txt)を使用した場合におけるソフトウェアの簡単な操作の流れを説明します。使用方法の詳細や各種クラスタリングの説明については「Cluster and TreeView」のマニュアルをご参照ください。

「Cluster and TreeView」は下記のアドレスよりダウンロードします。

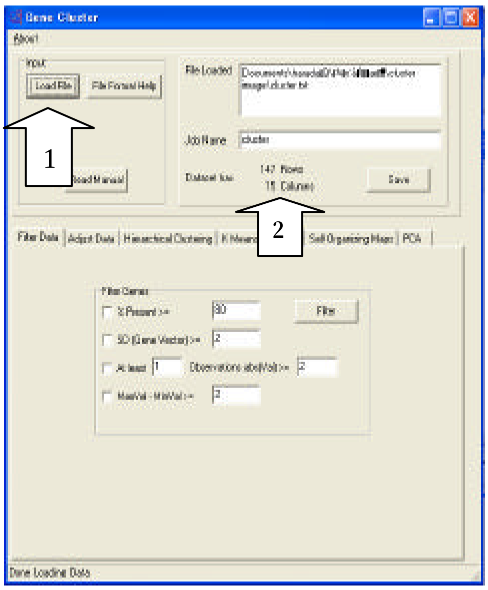
http://rana.lbl.gov/EisenSoftware.htm

＜Top page＞



Top page のsoftwareを選択し、Cluster と TreeView それぞれのソフトおよびマニュアルをダウンロードしてください。なお、本ソフトはwindows専用となっています。ダウンロードした自己解凍ファイルをクリックし、フォルダにあるSETUP .EXE アプリケーションファイルをダブルクリックし、インストールを行います。インストール終了後、インストール先のフォルダにはCluster.exeとTreeView.exeの二つのアイコンが作成されます。

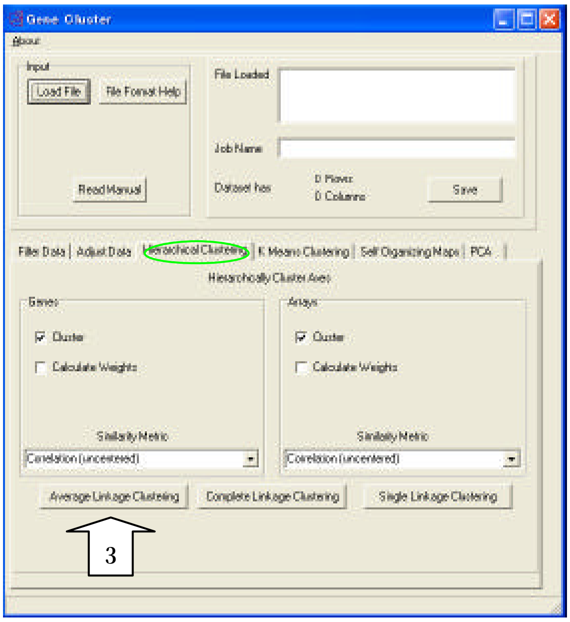
1. Cluster解析



①Clusterソフト を立ち上げると左図の初期画面が表示されます。

②矢印 1をクリックし、クラスタリングしたいファイルを読み込ませます。ここでは「Microarray Data Analysis Tool」で出力されたデータファイル(cluster.txt)を読み込ませます。「Microarray Data Analysis Tool」で出力されたデータファイルはすでに、log2比に変換されています。

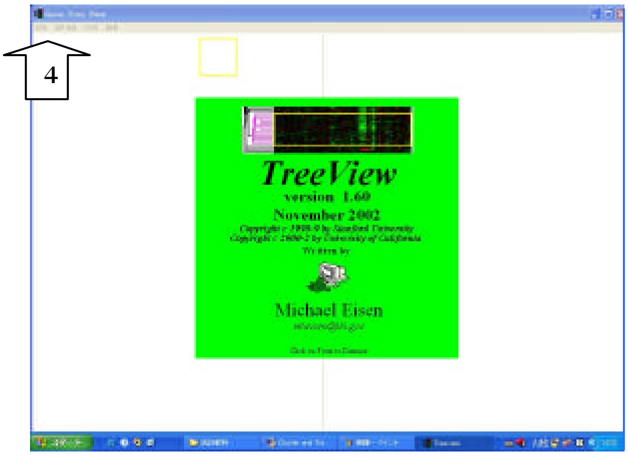
ファイルが適切に読み込まれると、読み込まれたデータの行数と列数が表示されます。



③Hierarchical Clusteringのパネルを選択します。
遺伝子（or 抗体)のクラスタリング(Geneの Clusterという部分をチェック)だけでなく、さまざまな組織で実験をしたときなどの組織間のクラスタリング（Arraysの Clusterという部分をチェック)も同時に実行することができます。

1,Average linkage clustering : 各クラスターの平均の位置間の距離を2つのクラスター間の距離とする方法。
2,Complete linkage clustering: 最遠位値にある2つの遺伝子（or 抗体)間の距離を2つのクラスター間の距離とする方法。
3,Single linkage clustering: 2つのクラスターの中で最近接位値にある2つ遺伝子（or 抗体)間の距離を2つクラスター間の距離とする方法。

④希望のクラスタリング法を選択し、実行ボタンを押します。ここでは、1,Average linkage clusteringを選択しています。
クラスタリング終了後、TreeView用の入力ファイルが（cdtファイルとgtrファイル）が自動作成されます。

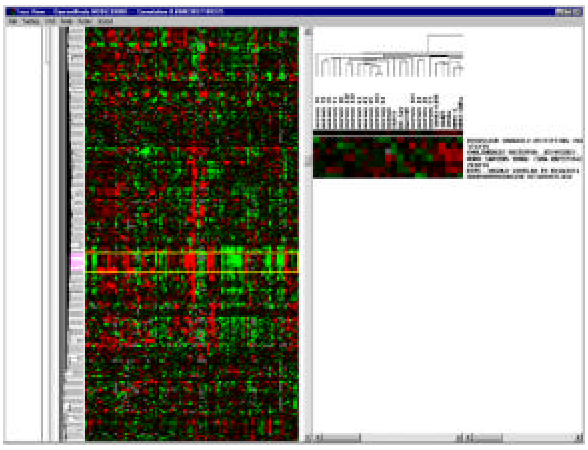


⑤TreeViewClusterソフトを立ち上げます。左図の初期画面が表示されます。

⑥矢印4をクリックし、用意したTreeViewの入力ファイル（cdtファイル)を読み込ませます。

⑦データ読込後、クラスターイメージが表示されます。サムネイルイメージの見たい部分をマウス操作で選択すると、右側の画面に拡大イメージが、6.クラスタリング用出力ファイルで選択した項目名とともに表示されます。

その他、詳細な使用方法については「Cluster and TreeView」のマニュアルをご参照ください。



12　よくあるご質問

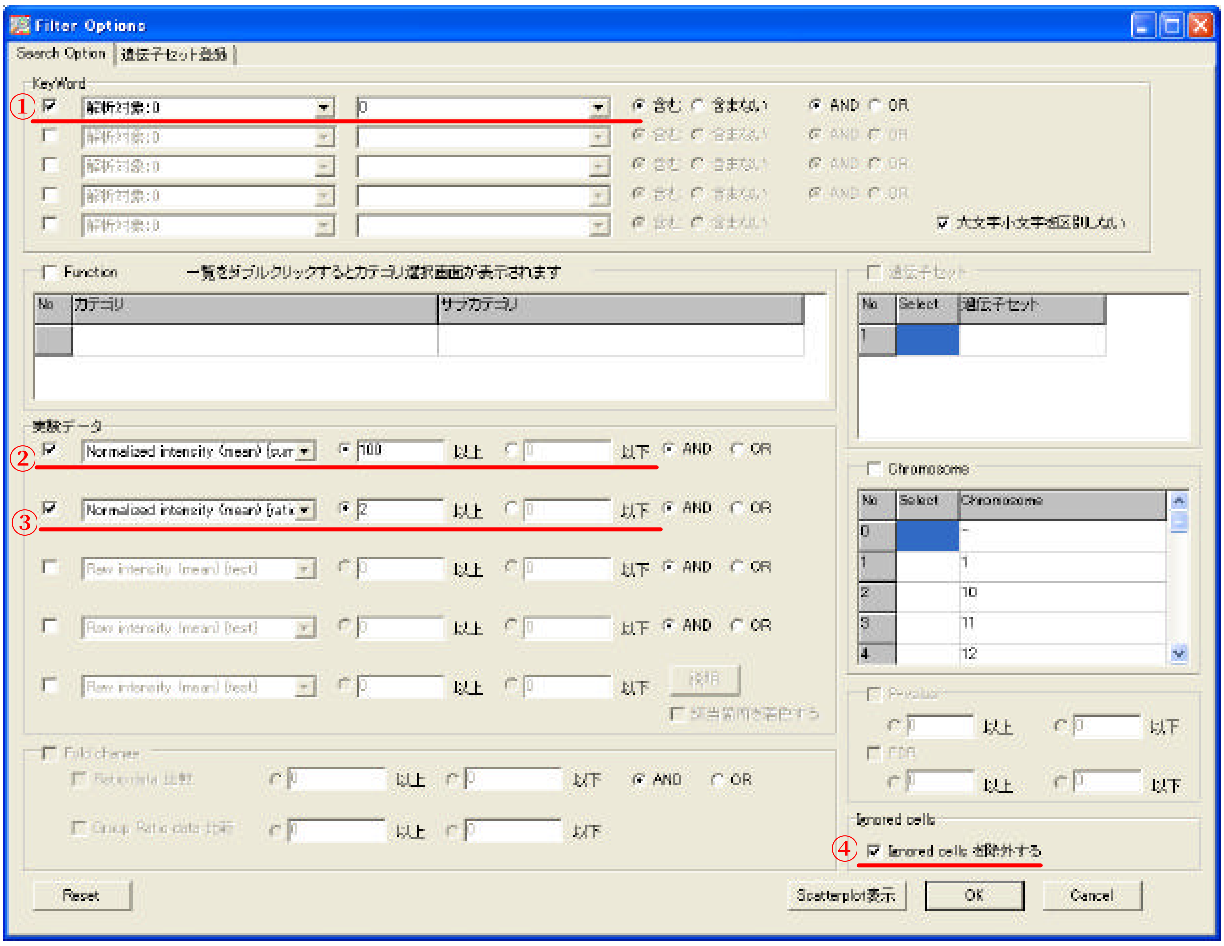
Q1. Filter Optionsで検索しても、遺伝子件数が納品Excelデータの「2Up」、「2Down」フラグのついた遺伝子数と同じ数にならない。(1比較解析)

A1. 納品用Excelデータでは、Fold Change（Ratio)が2倍以上増加、または減少している遺伝子に対してフラグをつけ、変動遺伝子を簡単に抽出できるようにしてあります。しかし本ソフトウェアで同様の抽出を行うには、Filter Options検索で様々なパラメータを設定しなければいけません。設定を行うパラメータは以下のとおりです。

①項目「Keyword」で「解析対象：0」に対して、「0」を含むとなるように設定する。(P.18参照)
②項目「実験データ」で「Normalized intensity (mean) {sum}」に対して、「納品Excelデータに記載されているCutOff値」以上となるように設定する。(P.19、26参照)
③項目「実験データ」で「Normalized intensity (mean) {ratio}」に対して、「2」以上または「0.5」以下となるように設定する。(P.19参照)
④項目「Ignored cells」で「Ignored cellsを除外する」にチェックを入れる。(P.20参照)

Ratioに関しては入力を忘れることは少ないのですが、CutOffの値や、特に解析対象とIgnored cellの設定を忘れている例がよく見られます。これらの設定をせずに検索を行うと、信頼性の低いノイズレベルのデータやコントロール遺伝子なども抽出されてしまいます。Filter Optionsでの検索時には、「解析対象：0」と「Ignored cells」の設定を忘れずに行ってください。

【Filter Optins設定例】



Q2. ノーマライゼーションについて、どの手法を選択すればよいか分からない。(Signal比較解析)

A2. Signal比較解析機能を使用すると、実験データの読み込みの過程で、「Per Array Normalization」と「Per Gene Normalization」の2種類のノーマライゼーションを実行する必要があります。それぞれ3～4つほどの手法の中から選択することになるのですが、どれを選択すればよいか分からないこともあると思いますので、ここで弊社で推奨している手法を紹介します。

弊社マイクロアレイ受託解析サービスでの、単色実験データに対するPer Array Normalizationには、多くの場合「Median」、または「Quantile」で補正を行っています。「Median」は単色のマイクロアレイデータの補正に最もよく使用されており、データの比較も可能になるのですが、実験日や実験環境の違いによってマイクロアレイ間に偏りができる場合があります。「Quantile」はそのような偏りを補正することができますが、発現プロファイルが極端に違うデータ同士に対して使用すると、それらの有意な違いも補正してしまうことがあります。よって最初は「Median」で補正を行い、Scatter Plotでマイクロアレイ間の偏りを確認し、偏りがあった場合のみ「Quantile」で補正し直すのがよいでしょう。

Per Array Normalizationに比べ、Per Gene Normalizationはどの手法を選んでも大きな違いはありません。しかし実験データの中に、ただ1つの明確なコントロールとなるものがない場合、Per Gene Normalizationでは「Median」を選択することが一般的です。対してコントロールとなる実験データがある場合は、「Control Experiment」で比較の基準とする実験データを選択します。

Q3. Gene OntologyまたはPathway解析で、出力したリスト内のどのデータが有意なのか分からない。

A3. 本ソフトウェアでは、各GO Termやパスウェイの該当遺伝子数などの解析データをリスト形式で出力することができますが、はじめはデータの量や種類の多さに戸惑うかもしれません。ここでは弊社で推奨している、リスト中からの有意なGO Termやパスウェイの抽出方法を紹介します。

本ソフトウェアからの出力リストには、各GO Termやパスウェイに対して「Changed Genes」、「Total Genes」、「Z-score」、「P-value」の4種類の数値データがあります。しかしデータの抽出には、これら4種類のうち2種類、「Z-score」と「P-value」のみを使用し、「Z-score>0であり、なおかつP-value<0.05」のGO Termやパスウェイを有意なものとしています。これら数値データの詳細はGene Ontology解析のページ (P.36)をご覧下さい。

Term	Changed Genes	Total Genes	Z score	P-value	Gene ID
cytoplasm	239(121)	5270(3279)	3.66	2.54E-7	ENSG00000002549,ENSG0000
ribosome	26(26)	233(233)	6.21	1.04E-6	ENSG00000037241,ENSG0000
electron transport	30(30)	420(420)	5.524	2.46E-6	ENSG00000051523,ENSG0000
generation of precursor metabolites and energy	32(2)	482(62)	5.24	5.29E-6	ENSG00000051523,ENSG0000
mitochondrion	55(55)	839(839)	4.614	2.19E-5	ENSG00000002549,ENSG0000
endoplasmic reticulum	43(43)	705(705)	3.58	0.00117	ENSG00000051108,ENSG0000
receptor activity	19(19)	1315(1315)	-2.849	0.00163	ENSG00000008300,ENSG0000
signal transducer activity	28(9)	1650(345)	-2.563	0.00417	ENSG00000008300,ENSG0000
organelle	263(0)	6684(0)	1.504	0.01122	ENSG00000002549,ENSG0000
RNA binding	24(24)	525(525)	2.592	0.01861	ENSG00000031698,ENSG0000
nucleus	131(129)	4273(4244)	-1.822	0.02008	ENSG00000004478,ENSG0000
electron carrier activity	8(8)	129(129)	2.422	0.02934	ENSG00000051523,ENSG0000

赤 :有意なGO Term　　青 :有意ではないGO Term

13. 履 歴

本説明書に関する変更の履歴となります。

- 2005 年 7 月 ver.1.0 取扱説明書の作成
- 2005 年 10 月 ver.1.1 取扱説明書の修正
- 2006 年 4 月 ver.1.2 取扱説明書の修正
- 2006 年 9 月 ver.1.4 取扱説明書の修正
- 2007 年 6 月 ver.2.0 取扱説明書の修正
- 2008 年 3 月 ver.3.0 取扱説明書の修正

Filgen®
biochip & nanofilm technologies